Modelling metabolism of contractile cells using constraint-based methods

Dobrin, Andrei

Awarding institution:
King's College London

Modelling metabolism of contractile cells using constraint-based methods

Andrei-Costin Dobrin

Centre for Human & Applied Physiological Sciences

Faculty of Life Sciences & Medicine

A thesis presented for the degree of Doctor of Philosophy at

King's College London

2020

Abstract

Constraint-based genome-scale modelling is a widely used approach that facilitates computational exploration of biological networks of many organisms and diseases. This modelling approach is an efficient alternative to wet lab experiments for analysing cell metabolism and physiological functions. A plethora of constraint-based methods have been developed to facilitate a comprehensive analysis of these models. These methods can be applied to generate testable computational predictions and hypotheses.

The work carried out in this thesis was focused on developing and analysing three genome scale constraint-based models for the normal fibroblasts (control), TGF-β-stimulated-fibroblasts (myofibroblasts) and skeletal muscle cell. These models can provide valuable insight into the cell metabolic pathways and even study the metabolism of certain diseases. Fibroblast and myofibroblast models were compared to investigate the effects of TGF-β on cell proliferation, glycolysis and collagen synthesis. Having compared both models, the work then focused further on the myofibroblast model and investigated the metabolic pathways leading to collagen synthesis. Genes and reactions which are essential to collagen synthesis were identified. Most of the results generated by simulations were shown to support *in vivo* observations and could pave the way for identifying drug targets for patients with Idiopathic Pulmonary Fibrosis.

In separate modelling, skeletal muscle simulations focused on studying the effects of hypoxia and amino acid supplementation on protein synthesis rate. This was with the view to providing valuable insights into metabolic requirements of muscle hypertrophy, possible causes (and treatments) of muscle atrophy, the cell responses to hypoxic stimulus and underlining the metabolic changes that occur in muscle during hypoxia. The modelling results showed that hypoxia directly impacts the protein synthesis rate and a metabolic switch from oxidative to glycolytic metabolism occurs. Furthermore, given the controversies surrounding amino acid supplements, studying the effects of amino acid supplementation on protein synthesis *in silico* can be an appropriate alternative to *in vivo* studies that aim at identifying the optimal combination of amino acids that could potentially trigger anabolic responses and increase protein synthesis in skeletal muscle. The results suggested that supplementation with certain combinations of amino acids could impact protein synthesis in muscle. Although further computational and experimental work may be required to verify these results, the current model predictions might pave the way for designing nutritional strategies that can positively influence the protein turnover in skeletal muscle.

Dedication

This thesis is dedicated to the memory of my grandparents, Anicuta and Gheorghe, for their love and unconditional support.

Acknowledgements

I would like to thank my primary supervisor, Dr. Lindsay Edwards, for giving me the opportunity to undertake this PhD project and for supporting me during these years. He believed in me since the first day we've met and gave me endless support. His unstoppable optimism, enthusiasm and dedication have made this thesis possible. I also owe a great deal of gratitude to my second supervisor, Professor Stephen Harridge, who was always there for me when I needed him. Without his patience, guidance and support I would have had a much harder time. Also, I want to thank Professor Ton Coolen who gave me useful advice at the early stages of the project.

I would also want to thank Dr. Thomas Pfau from University of Luxembourg for his directions and support. I learned a great deal from his senior experience in this area of research and I was lucky enough to learn from somebody like him. I would like to acknowledge the help from everyone in the Fibrosis DPU at GSK. In particular, I want to thank to Adam Taylor and Andy Blanchard for helping me understand the biological relevance of the fibroblasts, myofibroblasts and IPF. Also, Adam helped me understand the data I was using for my reconstructions and I want to thank him for his patience. I also want to thank everyone in the CHAPS department at King's who always provided support and advice on any issue.

Finally, and above all, I want to thank my loving parents Mariana and Traian who supported me through my entire academic journey and without whom this work would not have been possible.

List of abbreviations

3-PG – 3-Phosphoglycerate
AA – Amino Acid
ACAD9 – Acyl-CoA Dehydrogenase Family Member 9 (isobutyryl-CoA)
ACO1 – Aconitase 1 (cytoplasmic)
ACO2 – Aconitase 2 (mitochondrial)
ALDOA – Fructose-Bisphosphate Aldolase
AMP – Adenosine Monophosphate
ATP – Adenosine Triphosphate
ATP5F1B – ATP synthase
BCAA – Branched-Chain Amino Acids
BCKDHA – 2-Oxoisovalerate Dehydrogenase
CBM – Constraint-Based Model/Constraint-Based Modelling
COBRA – Constraint-Based Reconstruction and Analysis
COPD – Chronic Obstructive Pulmonary Disease
COX – Cytochrome *c* Oxidase
DEA – Differential Expression Analysis
DLST – Dihydrolipoyllysine-Residue Succinyltransferase
EAA – Essential Amino Acid
ECM – Extracellular Matrix
ENO1 – Enolase
FADH2 – Flavin Adenine Dinucleotide
FBA – Flux Balance Analysis
FB_N – Normal Fibroblasts (control)
FB_{TGF} – TGF-β-Stimulated Fibroblast
FH – Fumarase
FN – Fibronectin
FVA – Flux Variability Analysis
GAPDH – Glyceraldehyde-3-Phosphate Dehydrogenase
GDP – Guanosine Diphosphate
GEA – Gene Essentiality Analysis
GK1 – Guanylate Kinase 1
GLM – General Linear Model
GMP – Guanosine Monophosphate
GPI – Glucose-6-Phosphate Isomerase
GPR – Gene-Protein-Reaction Associations
GTP – Guanosine triphosphate
HIBADH – 3-Hydroxyisobutyrate Dehydrogenase
HIF-1 – Hypoxia-Inducible Factor 1

HK – Hexokinase
IDH2 – $NADP^+$-dependent Isocitrate Dehydrogenase
IDH3 – NAD^+-dependent Isocitrate Dehydrogenase
IPF – Idiopathic Pulmonary Fibrosis
LDHA – Lactate Dehydrogenase A
LDHB – Lactate Dehydrogenase B
LP – Linear Programming
MDH1 – Malate Dehydrogenase (cytosolic)
MDH2 – Malate Dehydrogenase (mitochondrial)
NAD – Nicotinamide Adenine Dinucleotide (oxidized form)
NADH – Nicotinamide Adenine Dinucleotide (reduced form)
NADP – Nicotinamide Adenine Dinucleotide Phosphate
NDUFA10 – NADH Dehydrogenase
NME – Nucleoside Diphosphate Kinase 1
OGDH – Alpha-Ketoglutarate Dehydrogenase
PC – Pyruvate Carboxylase
PCr – Phosphocreatine
PDH – Pyruvate Dehydrogenase
PDK1 – Pyruvate Dehydrogenase Kinase 1
PFKM – Phosphofructokinase
PGAM – Phosphoglycerate Mutase
PGK – Phosphoglycerate Kinase
PHGDH – Phosphoglycerate Dehydrogenase
Pi – Inorganic Phosphate
PKM – Pyruvate Kinase
PPi – Pyrophosphate
PSAT – Phosphoserine Aminotransferase
PSPH – Phosphoserine Phosphatase
PYCR1 – Pyrroline-5-Carboxylate Reductase 1
RNA-seq – RNA Sequencing
SHMT1 – Hydromethyltransferase (cytoplasmic)
SHMT2 – Hydromethyltransferase (mitochondria)
SL – Synthetically Lethal
SUCLG1 – Succinyl-CoA Ligase (GDP forming)
TCA – Tricarboxylic Acid Cycle
TGF-β – Transforming Growth Factor
TPI – Triosephosphate Isomerase
TPM – Transcripts Per Million
UQCR10 – Ubiquinol-6 Cytochrome *c* Reductase
α-SMA – Alpha-Smooth Muscle Actin

1 Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

The use of computational approaches in biology started in the 1960s with the advent of modern computers. These new approaches in biological studies were initially used for the computational interpretation of complex biochemical processes and to deal with increased amounts of experimental data (Hagen 2000). Nowadays, mathematical and computational methods are widely used tools in the study of biology. Systems biology is a multidisciplinary field of study that combines biological, mathematical and computational methods and techniques to provide a comprehensive exploration of complex biology systems (Friboulet et al. 2005).

There are many approaches to understanding complex biological systems. Rather than studying isolated parts or components of a biological system, systems biology transcends reductionist views and uses a holistic approach to provide a comprehensive analysis at all levels. This “system-level” approach allows an integrative analysis of biological functions and organisms by revealing interactions among components within a system whilst describing the “whole picture”. Recent technological advancements in computer science and genome sequencing techniques which have resulted in large and complex data sets have also facilitated and driven the use of this integrative approach (Westerhoff et al. 2004). Furthermore, vast amounts of “omics” data generated from novel technologies such as genomics, transcriptomics, proteomics and metabolomics have the potential to provide a comprehensive view of the functioning biological systems and their integration (Westerhoff et al. 2004). To cope with ever larger data sets, sophisticated computational tools are required, which has been facilitated by a dramatic increase in processing power.

Several computational approaches exist in the field of systems biology. Among them, genome-scale constraint-based modelling (CBM) is a widely used approach to interpret and analyse "omics" data in a biochemically constructive manner. Genome-scale models coupled with constraint-based analysis tools provide context for "omics" data and have multiple applications (Oberhardt et al. 2009) including model-driven hypothesis formulation (Ibarra et al. 2002), strategies for metabolic engineering (Kim et al. 2015), and *in silico* drug discovery (Folger et al. 2011). These models provide an efficient alternative to the study of metabolism in organisms or tissues and metabolic interactions associated with diseases (Vo et al. 2007, Bordel 2018). The overall aim of the work presented in this thesis is to use several constraint-based approaches to build and analyse two separate genome-scale metabolic models of two human contractile cells: myofibroblasts and skeletal muscle cells.

Myofibroblasts are contractile cells that are phenotypically situated between fibroblasts and smooth muscle cells and play a crucial role in the process of wound healing. Myofibroblast cells migrate to the sites of injury where they produce collagen and eventually contract the wound. Furthermore, their sustained presence at the site of injury and uncontrolled proliferation are suspected to be the main causes for fibrotic diseases.

Skeletal muscle cells are another essential contractile cell type, enabling movement, maintenance of body posture, breathing, etc. These cells are responsible for rapid and forceful contractions of short durations. Any loss of skeletal muscle mass or function is suspected to be determined by a lack of equilibrium between muscle protein synthesis and breakdown rates (McKinnell et al. 2004). Clinically, this can happen in conditions such as sarcopenia, chronic obstructive pulmonary disease (COPD) or cancer (Passey et

al. 2016). It can also occur in healthy people under environmental hypoxia. For instance, it was reported that loss of muscle mass can occur in mountain climbers (Boyer et al. 1984, Rose et al. 1988).

Despite some commonalities in their contractile phenotype, these cells have very different physiology and metabolic requirements. They are also both important to several phenotypes and diseases. The reconstruction and analysis of the genome-scale models for these two cells can produce meaningful insights into their metabolic capacities with therapeutic potentiality. Hopefully, model-driven predictions that are experimentally validated could be used for designing optimal patterns of nutrient supply or identify candidate drug targets.

Thus, the work presented here seeks to use constraint-based modelling approaches to reconstruct and analyse the genome-scale metabolic networks for these contractile cells.

1.2 Literature review

This literature review aims to consider a number of areas that are pertinent to genome-scale CBM and a general description of the organisms that are reconstructed using constraint-based methods. It will give a brief review of metabolism, metabolic pathways, metabolic network reconstruction, systems biology and CBM approaches. Lastly, this section provides a description of the biological relevance of the fibroblasts, myofibroblasts and skeletal muscle cells, the cells types which are the focus of the modelling work to be presented in this thesis.

1.2.1 Metabolism

Metabolism is the set of chemical reactions that produce energy and maintain life in an organism. Metabolism can be divided into two overarching processes: catabolism and anabolism. Catabolism is the process of degrading larger molecules into smaller ones. This process generates energy which is ultimately stored in the form of adenosine triphosphate (ATP), an exergonic process. On the other hand, anabolism uses energy to build up more complex molecules from simple precursors, an endergonic process. Chemical reactions are usually catalysed by enzymes and these play a crucial role in both catabolism and anabolism. Enzymes are proteins that lower the activation energy for certain reactions, and thus increase the rate at which these reactions occur (Cooper 2000). Without the intervention of enzymes, many reactions would happen at very low rates and life could not be maintained (Wolfenden et al. 2001).

Morphologically, enzymes are composed of large chains of amino acids (AAs) which fold up into a tertiary structure. During the process of translation, the information contained in the messenger ribonucleic acid (mRNA) is translated into a sequence of AAs (Kuchel 1998) to form proteins. This is preceded by transcription: the copying of genetic information from DNA sequence to mRNA. In eukaryotic cells these processes occur one after the other and in different places within the cell: Transcription occurs in the nucleus, while translation takes place in the cytoplasm. On the other hand, in prokaryotic cells, these processes are closely connected and happen in the same cell compartment. Furthermore, in prokaryotic cells translation can begin while transcription is still happening (Berg et al. 2002). Taken together (transcription and translation), this two-step process, is also known as the central dogma in molecular biology (Crick 1970) and it describes a directional flow of information from DNA to RNA and from RNA to proteins.

1.2.1.1 Metabolic pathways

In order to convert one metabolite into another, a series of enzyme-catalysed reactions may be required. Metabolic pathways are useful tools for illustrating this process. In this framework, metabolic pathways can be defined as a sequence of successive reactions that connect metabolites (Kremling 2013). Metabolic pathways can be linear, branched (convergent or divergent) cyclic or spiral (Fig. 1.1).

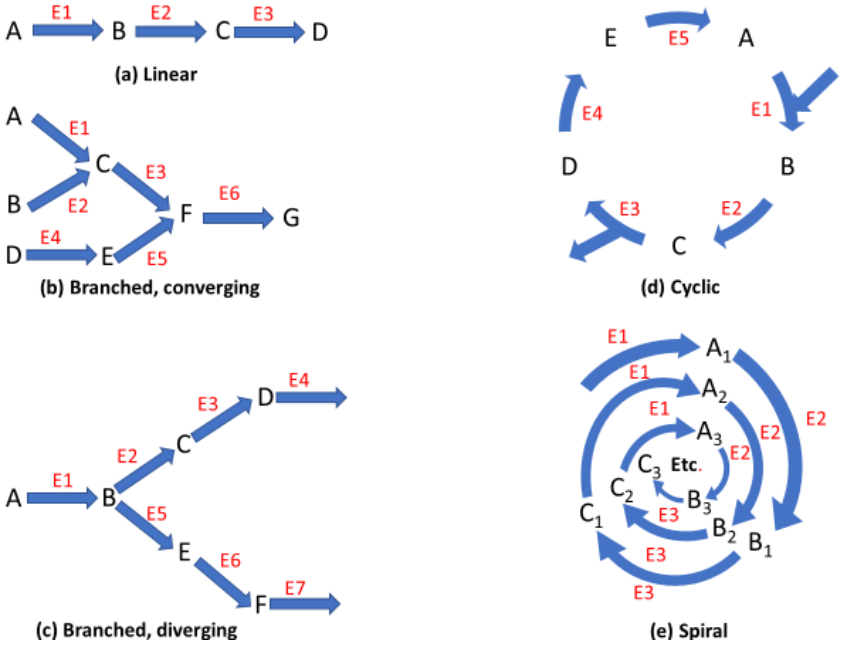


Figure 1.1 Patterns of metabolic pathways
Metabolic pathways can be linear, cyclic, branched or spiral. In linear metabolic pathways, a compound is transformed into another through a series of intermediate reactions. Branched pathways can be either convergent (several compounds combine to produce a single compound further along in the pathway) or divergent (an intermediate compound can yield several end products). In cyclic pathways, the final compound is the same as initial substrate and intermediates are regenerated at every step. In spiral pathways, the same enzymes catalyse a series of reactions that are used to build up or break down a certain compound. Adopted from (Boyer 2001) Copyright ©2002 Wadsworth Group, a division of Thomson Learning, Inc.

Primary metabolism consists of several core reactions that are necessary for maintaining life. Thus, there are a number of metabolic pathways that are common among most organisms. Examples of such pathways include glycolysis, tricarboxylic acid (TCA) cycle or pentose phosphate pathway.

Metabolic pathway analysis allows an accurate investigation of the structural and functional properties of the metabolic network (Schilling et al. 1999) and can be used to identify meaningful routes within the network (Rugen et al. 2012). With the advent of network reconstructions and genome-scale metabolic models, metabolic pathway analysis can be integrated with constraint-based approaches such as Flux Balance Analysis (FBA) (Orth et al. 2010) to assess the impact of various genetic and environmental perturbation on the entire network (Schilling et al. 2000). These approaches have enabled a better understanding of metabolic capabilities and limitations for several organisms (Schilling et al. 2000) and will be described in detail in sections 1.2.3-1.2.4.

1.2.2 Systems biology concepts

Systems biology is a multidisciplinary field that combines mathematical, computational and experimental approaches to explore various biological systems (Edwards 2017). At the core of this field is the use of holistic instead of reductionist concepts when studying a complex biological system. A system can be defined as a collection of interacting elements that form an integrated whole. A reductionist view aims to divide a system into constituent parts and study them separately. Its historical roots can be traced back to the late 17th century when many scientific discoveries were based on this approach (Trewavas 2006). On the other hand, a holistic view describes systems "as a whole" by considering the interactions between the constituent components. In general system theory, a holistic approach is closely related to the concept of emergence (von Bertalanffy 2003). A system's emergent properties do not belong to any individual part of the system but arise from the system as a whole (Kesić 2016). Thus, systems biology

aims to study an organism as an integrated system of interacting components rather than focusing on individual parts.

Developments in molecular biology have led to a transition from a reductionist view on biological systems, to a more integrative approach (Ge et al. 2003). Scientists became able to study the interactions that define cellular processes. Furthermore, the advent of genome-sequencing techniques contributed highly to this transition (Westerhoff et al. 2004). High-throughput “omics” technologies such as genomics, transcriptomics, proteomics and metabolomics allow the combined analysis of several classes of molecule such as DNA, RNA, proteins and metabolites. These highly-parallel technologies permit the simultaneous interrogation of different molecules of the same cell (Fischer 2008) and provide a comprehensive view of biological systems (Robinson et al. 2016). A system-level view of biological systems can bridge the genotype-phenotype gap.

One of the most important challenges in biology is understanding the genotype-phenotype relationship. This relationship describes the impact of variations in genome on various functional phenotypes. This is a complicated concept, and although some attempts have been made to formalise this relationship and characterize the two biological components simultaneously (Alberch 1991), not many consistent results have been obtained. This difficulty could be attributed to the complexity of biological systems. However, technological advancements in genome sequencing techniques and the generation of genome-wide data have now facilitated the development of biological networks which provide a mechanistic framework for encoding the genotype-phenotype relationship (Lewis et al. 2012).

With the increased number of large datasets containing genome-wide information, advanced mathematical and computational tools are needed for processing, integrating and analysing the data. Biological networks provide a platform for integrating "omics" data and facilitate computational investigation of the biological systems. These networks can successfully replicate a biological system of interacting components. Furthermore, biological networks can be converted into mathematical models that allow computational analysis and *in silico* simulations. Many types of molecular networks can be distinguished: metabolic, protein-protein interaction, signalling and gene regulatory networks (Mahadevan et al. 2005). Among all types of biological networks, metabolic networks are the most widely studied due to their unique properties and applications (Pfeiffer et al. 2005, Edwards 2017). These properties and applications will be described in the next section.

1.2.3 Metabolic networks

Metabolism can be described by a complex network of chemical processes. Metabolic networks consist of metabolites, chemical reactions that produce metabolites, enzymes that catalyse reactions and genes that encode various enzymes. Thus, a metabolic network is essentially an interconnected network of metabolites, reactions and enzyme-coding genes. Metabolic networks can be represented by a directed graph where nodes are represented by metabolites, and edges represent reactions (Fig. 1.2). The structure is different from signalling or protein-interaction networks where substrates and enzymes are not separated, and edges are determined by interactions and describe whether an interaction is present or absent. The separation between reactions and metabolites in the metabolic networks allows the conservation of mass or energy at every node by imposing physicochemical constraints such as the mass or energy balance

equations. Thus, the unique properties of metabolic networks are determined by their structure and are described by the separation between enzymes and substrates and the conservation constraints that could be imposed at every node (Mahadevan et al. 2005).

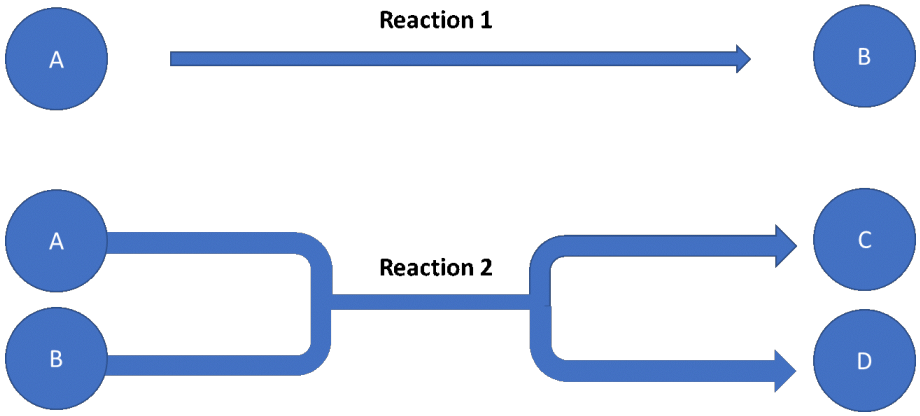


Figure 1.2 Metabolic pathways in the form of directed graph
Metabolic networks in the form of directed graphs comprised of edges to represent reactions and nodes to represent metabolites.

Genome-scale metabolic networks are typically converted into a mathematical model that facilitates the computational investigation. Once a metabolic network is converted into mathematical model, its functional states (phenotypic functions) can be investigated using graph methods or constraint-based approaches such as FBA (Orth et al. 2010). Details about network reconstruction process and modelling approaches are further discussed in sections 1.2.3.1-1.2.3.2 and 1.2.4.

1.2.3.1 Metabolic network reconstruction

A prerequisite for developing a model of a target organism is the reconstruction of a genome-scale metabolic network. There are essentially two approaches in the reconstruction process: inference-based and knowledge-based (Hyduke et al. 2013, Katze 2013). Inference-based reconstruction uses statistical approaches to estimate and infer unknown interactions in the network, based on interaction patterns in high-throughput data. However, there are certain limitations of the inference methods

(Margolin et al. 2007). For instance, the inference network problem is underdetermined, or, in other words, it has an infinite number of solutions. Thus, many solutions can match the available experimental data, but not all of them can be biologically relevant (De Smet et al. 2010).

Knowledge-based reconstructions are built using curated genomic and metabolic data that are based on the annotated genome and experimental information from literature. In other words, metabolic network reconstructions detail the known metabolic reactions in an organism and the genes that encode each enzyme. The genotype-phenotype relationship is described by Gene – Protein – Reaction (GPR) associations that match genes to enzymes and enzyme-catalysed reactions (Fig.1.3). Hence, genetic data is mapped to flux data and one can assess the effect of genetic perturbations on various phenotypes. In a metabolic network reconstruction, Boolean logic is used to represent the GPR associations (Fig. 1.4).

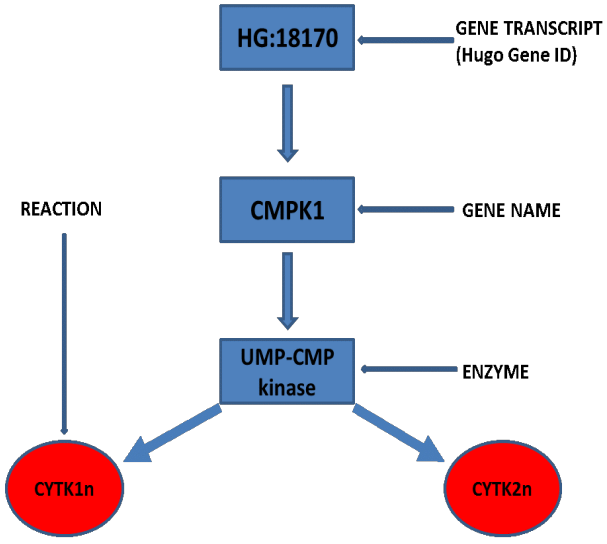


Figure 1.3 Gene-protein-reaction (GPR) relationship
An example of GPR relationship from a genome-scale-metabolic network. Genes are matched to enzymes and reactions.

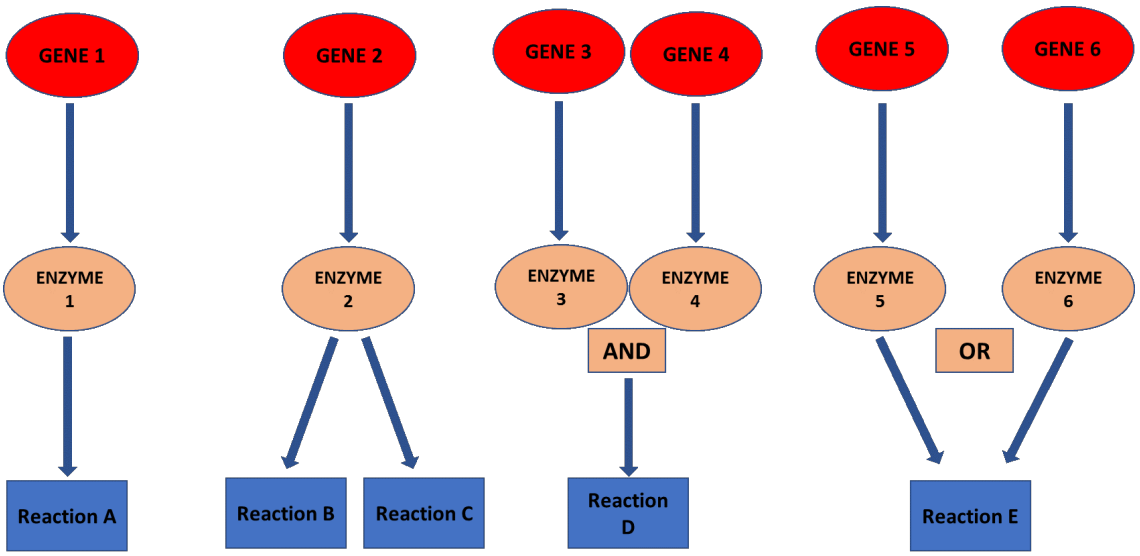


Figure 1.4 GPR and Boolean logic
Boolean operators (AND, OR) can be used to describe the GPR associations. Adapted from Jensen et al. (2011)

Reconstruction of metabolic networks is a complex, laborious and iterative process. In 2010 a protocol for metabolic network reconstructions was published (Thiele et al. 2010) and divided the entire process into 96 steps and 4 stages. Furthermore, the time for constructing a comprehensive metabolic network was estimated to last from 6 to 24 months.

Genome-scale metabolic reconstructions have been extensively used for studying various organisms from prokaryotic cells to eukaryotic ones. Since the first metabolic reconstruction in 1999 (Edwards et al. 1999) various cells and tissues have been reconstructed (Forster et al. 2003, Borodina et al. 2005, Beste et al. 2007, Jamshidi et al. 2007, Oh et al. 2007). Once assembled, the reconstruction can be easily converted into a mathematical format (*in silico* model) that facilitates the computational investigation of several biological functions.

The first genome-scale reconstruction of the global human metabolic network (Recon 1) was published in 2007 (Duarte et al. 2007). In the same year, Edinburgh Human Metabolic Network (Ma et al. 2007), a second global reconstruction of the human

metabolic network was published too. Recon 1 contains physiological and biochemical information for 1,496 genes, 2,004 proteins, 2,766 metabolites, and 3,311 reactions. More than 1500 scientific publications were used to build Recon 1. Furthermore, this network reconstruction was transformed into a computational format (*in silico* model) and the simulation of 288 metabolic functions of different cells or tissues was used to validate this human metabolism model (Duarte et al. 2007). Since the first reconstruction in 2007, several updates have been published and they were incorporated in different versions of the global human metabolic reconstruction. Five detailed and complete reconstructions of the human metabolism are available to date: Recon 1 (Duarte et al. 2007), Recon 2 (Thiele et al. 2013), Recon 2.1 (Smallbone, K. 2013), Recon 2.2 (Swainston et al. 2016) and Recon 3D (Brunk et al. 2018).

Recon 2, published in 2013, was an extension of Recon 1 with significant improvements in terms of information coverage and functional properties. Recon 2 was more comprehensive than Recon 1, accounting for 1789 genes, 7740 reactions and 2626 unique metabolites. Recon 2 also reduced the number of blocked reactions (827 in Recon1 compared to 441 in Recon 2) and dead-end metabolites. In this framework, dead-end metabolites represented those metabolites which were either produced or consumed in only one reaction. Furthermore, a full set of metabolic tasks (metabolic functions such as biomass, ATP production etc) were fulfilled by Recon 2, while Recon 1 only managed to fulfil 63% of them. Apart from the improved metabolic coverage, Recon 2 also displayed advanced predictive capabilities when compared to Recon 1. Recon 2 successfully predicted more biomarkers for several inborn errors of metabolism (IEMs) and with higher accuracy when compared to Recon 1. In brief, the concept of predicting metabolite biomarkers for IEMs using metabolic networks refers to the analysis of the

exchange reaction fluxes and their comparison with the metabolite levels in the experimental data (Shlomi et al. 2009). Recon 2.1, an updated version of Recon 2, resolved several unbalanced reactions present in the previous reconstruction. These unbalanced reactions could lead to growth being simulated without any carbon source. However, Recon 2.1 couldn't resolve the mass and charge imbalances for every reaction in the reconstruction. Therefore, there was a need for a more developed reconstruction to amend these issues. Recon 2.2 integrated the Recon 2 updates that were previously published, ensured the mass and charge balance for all reactions and provided a better representation of the energy metabolism. ATP synthesis incorporated carbon availability and fatty acid oxidation (both mitochondrial and peroxisomal) was expanded. A new compartment, mitochondrial intermembrane space, was created to incorporate a proton gradient. In the previous reconstructions, protons were only represented as cytosolic protons. Hence reactions in the respiratory chain and aspartate-glutamate shuttle which involved proton transportation were updated with intramembrane protons. A combination of manual and semi-automatic curation was performed, leading to a comprehensive model consisting of 7785 reactions, 1675 genes and 5324 metabolites. Recon 2.2, the main reconstruction used for the modelling work undertaken in this thesis, was the most up to date reconstruction available when this project began. Since then, a more developed reconstruction has been published: Recon 3D, the latest version of the global human reconstruction, was published in 2019 and accounted for 3288 genes, 13543 reactions, 4140 metabolites and 12890 protein structures. This makes Recon 3D the most comprehensive metabolic human reconstruction to date. Recon 3D also provides the three-dimensional structure of proteins and metabolites which allows an integrative analysis of metabolic responses to drugs.

By using either manual or automated approaches, these global metabolic models can be coupled with high-throughput “omics” data such as transcriptomics, metabolomics or proteomics, to generate tissue and cell-specific models (Agren et al. 2012, Wang et al. 2012). For example, gene expression data for a certain cell or tissue can be integrated with the global metabolic reconstruction to determine cell-specific reactions. Coupling “omics” data with generic metabolic models has led to the development of several cell or tissue specific models (Bordbar et al. 2010, Chang et al. 2010, Gille et al. 2010, Jerby et al. 2010) and cancer specific models (Folger et al. 2011, Frezza et al. 2011, Jerby et al. 2012, Nilsson et al. 2017). The most common approach involves a combination of manual and automated methods and several algorithms have been developed for this purpose: GIMME (Becker et al. 2008), MBA (Jerby et al. 2010) , FASTCORE (Vlassis et al. 2014) or FASTCORMICS (Pacheco et al. 2015).

The GIMME algorithm initially considers a gene expression data set, a global metabolic reconstruction and certain biological functions as the main inputs. Reactions corresponding to a gene expression below a predefined threshold are declared inactive and removed from the network. The algorithm then optimizes for a specific biological function (objective function) such as ATP production or growth, etc. If the network fails to fulfill this required biological function, some reactions that were previously removed are reactivated subject to a minimal inconsistency with the gene expression data.

MBA considers a core set of highly active reactions, a second core set of medium confidence reactions, and tries to find a feasible network which contains the entire first set, a maximum number of the second set (if a specific trade-off is met) and a minimal number of reactions from the global human reconstruction. MBA is essentially pruning all reactions from the human reconstruction which do not support those reactions in the

first and second set. It is important to point out that MBA has very high computational demands.

Similar to MBA, the FASTCORE algorithm considers a core set of reactions and aims to use a minimum number of reactions from the global human reconstruction to create a context-specific model while trying to include all reactions from the core in the final model. Unlike MBA which uses pruning, the search for a minimal solution is done using the L1-normalisation, minimising the L1 norm of the vector of fluxes in the non-core set. The non-core set is essentially included in a "penalty" set and L1 normalisation is used to minimise fluxes in this set. This method significantly increases the speed of the algorithm.

FASTCORMICS is essentially a workflow that uses FASTCORE algorithm for model generation with a different discretization for gene expression and considers a "non-penalty set" for certain reactions. It is also important to keep in mind, that one idea of FASTCORE is to not add unnecessary reactions, and generate a small model of highly active metabolism, rather than a large one, where some reactions would need to be restricted to very low bounds because there is almost no enzyme expression to catalyse them. The idea of FASTCORMICS is to get rid of those reactions, which would be included in the model just because they are: i) catalysed by promiscuous genes, ii) transporters, iii) part of cyclic pathways, or iv) involved in shorter pathways for producing metabolite B from metabolite A despite the contradicting information from omics data. FASTCORMICS does not remove these reactions directly but assigns them to a "non-penalty set" which is removed from the core set but favoured over the the non-core set. FASTCORMICS is the workflow chosen for the modelling work in this thesis. Thus, a more

detailed description of the FASTCORE algorithm and FASTCORMICS workflow is included in the Methods chapter, section 2.2.4.

FASTCORE is a very restrictive algorithm and it will only allow reactions which are necessary for the use of the core to stay in the model. This is similar to the concept that MBA uses, while GIMME only removes reactions if they are actively unsupported, and relies a lot on the objective function selected. With GIMME, consistency is defined by the feasability of the objective function. This means that reactions which might not be relevant for a certain organism may be wrongly added to the model. FASTCORE does not require an objective function in the form of a biological function and the only mathematical objective is network consistency. The result of FASTCORE algorithm is not a flux distribution, but a set of reactions which are required so that all core reactions are able to carry a minimal flux. In the end, every algorithm has its advantages, but it is the author's opinion that FASTCORE/FASTCORMICS provides a solid basis for further curation in comparison with most of other algorithms in which retained reactions could also be unrelated to the organism being studied.

1.2.3.2 Conversion of metabolic network reconstruction to a computational format

Before a metabolic network can be investigated with computational approaches, it must be converted into a mathematical format. The stoichiometric matrix is the mathematical representation of the metabolic network reconstruction and the central component of genome-scale metabolic models. In this framework, the metabolic network is represented by a stoichiometric matrix $\boldsymbol{S_{ab}}$ with $\boldsymbol{a}$ metabolites and $\boldsymbol{b}$ reactions. The dimension of the matrix is $\boldsymbol{a\ x\ b}$ where $\boldsymbol{a}$ is the number is the number of rows and $\boldsymbol{b}$ is the number of columns. In the matrix $\boldsymbol{S_{ab}}$ each entry $\boldsymbol{s_{ab}}$ represents the stoichiometric coefficient of metabolite $\boldsymbol{a}$ in reaction $\boldsymbol{b}$**.** All the reactions in the matrix are represented

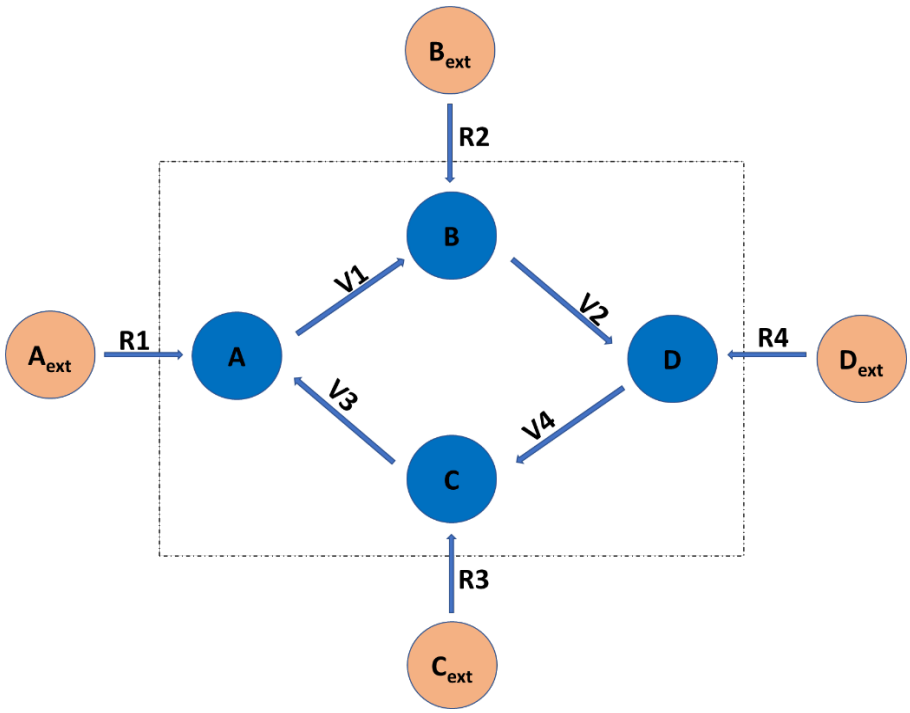
by the flux vector $\boldsymbol{v_b}$ which has the dimension equal to the number of reactions. The following simple ordinary differential equation represents the rate of change in the concentration of a metabolite over time:

$$\frac{\boldsymbol{dX}}{\boldsymbol{dT}} = \boldsymbol{Sv} \ (\mathbf{1.1})$$

where $\boldsymbol{X}$ is the vector of metabolite concentrations with dimension $\boldsymbol{a}$, $\boldsymbol{S}$ is the stoichiometric matrix with dimension $\boldsymbol{a\ x\ b}$, and $\boldsymbol{v}$ the flux vector with dimension $\boldsymbol{b}$.

The rate of change in the metabolite concentration is usually dependent on kinetic parameters. However, data regarding these time-variant parameters is scarce and their measurement is challenging (O'Brien et al. 2015). The assumption of steady state bypasses this obstacle and allows an investigation of the large networks based solely on reaction stoichiometry at steady states. For any internal metabolite, consumption and production rates are equal and the stoichiometric matrix can be divided into two sections (Fig. 1.5).

The steady-state assumption under mass conservation law is usually written as a mass balance equation and represents a constraint imposed on the genome-scale network. The imposition of further constraints constitutes the basis for CBM. This modelling approach is formally known as constraint-based reconstruction and analysis (COBRA).



	$R1$	$R2$	$R3$	$R4$	$V1$	$V2$	$V3$	$V4$
$Aext$	-1	0	0	0	0	0	0	0
$Bext$	0	-1	0	0	0	0	0	0
$Cext$	0	0	-1	0	0	0	0	0
$Dext$	0	0	0	-1	0	0	0	0
A	1	0	0	0	-1	0	1	0
B	0	1	0	0	1	-1	0	0
C	0	0	1	0	0	0	-1	1
D	0	0	0	1	0	1	0	-1

Figure 1.5 A metabolic network in steady state and its associated stoichiometric matrix A_{ext}, B_{ext}, C_{ext} and D_{ext} represent the external metabolites, while A, B, C and D represent the internal metabolites. R1, R2, R3, R4 represent the boundary fluxes while V1, V2, V3, V4 represent the internal fluxes.

1.2.4 Constraint-based modelling

1.2.4.1 Introduction to constraint-based modelling

According to Charles Darwin, organisms exist in a resource-limited environment and natural selection favours those who adapt to this environment (Darwin, 1859). Environmental pressure and resource allocation trade-offs shaped cellular behaviour during evolution. To optimise their resources and limit possible phenotypes, cells must

impose several constraints on themselves (Covert et al. 2003). Thus, constraints are essential for limiting cellular behaviour and defining achievable phenotypes (Palsson et al. 2015). When applied to reconstructed biological networks such as metabolic networks, a constraint-based approach can be useful for analysing the feasible functional states of the network, given a set of constraints (Bordbar et al. 2014). This approach leads to the computational investigation of reconstructed networks by using several constraint-based methods and techniques.

1.2.4.2 Constraint-based approaches

In a genome-scale metabolic network, the assumption of steady state implies that the rate of change for metabolites remains constant over time or, in other words, the rate at which metabolites that are produced is equal to the rate at which metabolites are consumed. Thus $\frac{\boldsymbol{dx}}{\boldsymbol{dt}} = \mathbf{0}$ and equation 1.1 can be written as:

$$\boldsymbol{Sv} = \mathbf{0} \ (\mathbf{1.2})$$

It is important to point out that, in a genome-scale model, the number of reactions or fluxes is greater than the number of metabolites. Thus, equation 1.2 can be written as a system of $\boldsymbol{a}$ equations with $\boldsymbol{b}$ unknown variables. This system of linear equations has no unique solution and all possible solutions are said to be in the null space of matrix **S** (Orth et al. 2010). The null space represents the solution space for the flux vector $\boldsymbol{v}$. In reality, not all solutions can be achieved as organisms are limited by certain constraints (Price et al. 2004). Therefore, additional constraints can be imposed to reduce the size of the solution space and limit the range of possible cellular behaviours. In this framework, several types of constraints can be distinguished: physicochemical, environmental, topological and gene regulatory (Price et al. 2004). From another perspective, constraints can be classified into “adjustable” and “non-adjustable”

constraints, based on their ability to adapt to evolutionary changes (Palsson 2000). Non-adjustable constraints are based on flux capacity, network topology and reaction reversibility (Covert et al. 2003). Physicochemical constraints fall in this category and they are associated with most genome-scale models. Adjustable constraints, such as gene regulatory ones, are used in second-generation models to further limit the solution space and describe the evolutionary cell behaviours (Covert et al. 2003, Palsson 2015). Cells receive signals from the extra- or intracellular environment, which determines changes in the expression of genes. In this way, cells adapt to environmental changes and these regulatory events may underlie optimal evolutionary behaviour. Gene regulatory constraints used in second-generation models are self-imposed by the cell, time-variant and dependent on certain biological circumstances (Covert et al. 2003).

Most commonly, physicochemical constraints are mathematically represented in the form of balances and bounds (Orth et al. 2010). Mathematically, balances are represented by equations, while bounds are imposed in the form of inequalities. For instance, equation 1.2 is called the mass balance constraint. Upper and lower bounds can be set to limit the flux capacity or change the reversibility of certain reactions:

$l_i \leq v_i \leq u_i \, \forall \, i \in N$ where v_i represents the flux vector, l_i and u_i represent the lower and upper bound respectively and N represents the set of natural numbers.

These bounds constrain the uptake or production rates for certain compounds. The values for rates are usually collected from the literature and they are based on experimental information. Constraints limit the solution space and the system of linear equations derived from equation 1.2 can be solved using linear programming (LP). A LP problem can be defined as the process of optimising a linear function given a set of constraints. Optimisation typically means maximisation or minimisation of a linear

function. Constraints usually take the form of linear equalities of inequalities and the linear function is often called the objective function. Here is an example for a typical LP problem:

Determine a_1 and a_2 to maximise $2a_1 + a_2$ given the following constraints:

$$a_1 + 2a_2 \leq 8$$

$$2a_1 + 4a_2 \leq 14$$

This simple example can be generalised. The generalisation is known as the General Standard Linear-Programming Problem (Bertsimas et al. 1997, Gass 2003).

Given n variables and m constraints:

$$maximize \sum_{j=1}^{n} c_j a_{ij} \; subject \; to:$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_j$$

$$i = 1,2, \dots, m$$

$$and$$

$$\boldsymbol{x_j \geq 0} \quad j = 1,2, \dots, n$$

A standard LP problem can be solved by using the simplex method. Dantzig's simplex method (Dantzig et al. 1987) is an algorithm for solving linear problems. It can be implemented computationally, and has a wide range of applications (Gass 2003) including FBA (Orth et al. 2010). Thus, once bounds and balances are imposed and the solution space is defined (Fig. 1.6), FBA can be employed to solve the following LP problem:

$$maximise\ or\ minimise\ \ Z_{obj} = \ c^t v$$

$$Subject\ to:$$

$$Sv \ = \ 0$$

$$l_i \leq \ v_i \leq \ u_i \ (LP\ 1.1)$$

Where v_i represents the flux for reaction i, c represents the row vector of weighting coefficients for fluxes in v. The objective function Z_{obj} is written as a linear combination of fluxes i.e. $Z = \sum_i c_i v_i$. This LP problem can be solved computationally by using existing algorithms and packages. These packages are further discussed in Chapter 2.

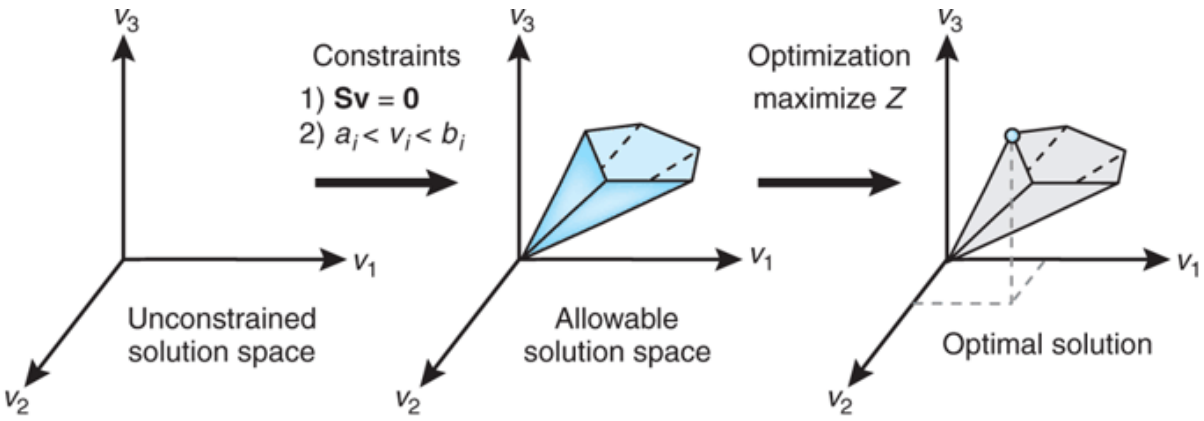


Figure 1.6 Constraints and Flux Balance Analysis
The steady-state constraint is imposed in the form of a mass balance equation. The imposition of upper and lower bounds for each reaction limits the solution space and FBA can be employed to find an optimal solution v to maximise Z within the bounded solution space. Adopted from Orth et al. (2010).

FBA seeks to optimise a certain biological function (objective function) given a set of metabolic constraints. These objective functions can include biomass or cell growth (Edwards et al. 2000, Forster et al. 2003, Feist et al. 2007), ATP production (Ramakrishna et al. 2001, Vo et al. 2004), or nutrient uptake (Schwender 2008). The most common objective function is the biomass yield which simulates growth of an organism. Biomass or cell growth is usually associated with microorganisms as they have a high proliferation rate (Agren et al. 2012). However, a biomass objective function was used to simulate growth/proliferation for other cells too, e.g. cancer cells (Bordel 2018), macrophages

(Bordbar et al. 2010) or fibroblasts (Vo et al. 2007). A biomass objective function takes the form of a stoichiometrically balanced reaction which contains the energy requirements and building blocks that are essential for an organism to grow.

1.2.4.3 Constraint-based methods

FBA (Orth et al. 2010) is a common approach used in CBM. When FBA is employed and biomass maximised or minimised, one can analyse the biosynthesis and conversion of certain biomass compounds that are required for optimising this function. Optimising cell growth using a biomass objective function has produced results consistent with experiments. For instance, in a study published in 2001, the biomass objective function was maximised and the results suggested that *Escherichia coli* uses its metabolism to reach maximal growth which is consistent with experimental data (Edwards et al. 2001). Furthermore, in another study which used biomass to simulate growth, the modelling work predicted the substrates along with the reactions and genes required for the growth of *H. pylori* with significant accuracy (Schilling et al. 2002).

One of the most important FBA-related applications is Gene Essentiality Analysis (GEA) (Edwards et al. 2000). This promising tool is FBA-based and uses computational approaches to study *in silico* genetic perturbations. GEA performs single-gene deletion to assess the essentiality of certain genes to biomass yield or other cellular functions, depending on the organism being analysed. FBA-based GEA produced results that were consistent with experimental data such as the identification of the essential genes for growth (Edwards et al. 2000). It is important to point out that FBA assumes an optimal growth. Thus, FBA-based GEA generates accurate predictions for wild-type strains but it is less accurate for genetically or environmentally perturbed strains that typically display sub-optimal growth (Segrè et al. 2002). MOMA (Segrè et al. 2002) was developed to

predict mutant fluxes by finding an approximate sub-optimal flux distribution, while minimising the difference between the wild-type and mutant optimum. Synthetic lethality analysis is performed in a similar way to assess synthetic interactions of genes. Synthetic Lethal (SL) genes were initially defined as those genes whose simultaneous knockout is lethal or prevents growth of an organism (Novick et al. 1989, Guarente 1993). Pairs (or more) of non-essential genes can be simultaneously deleted to assess their effect on growth rate or other cellular functions. Although some genes may be identified as dispensable in GEA, the combined knockdown of two non-essential genes could be lethal to the chosen objective function.

By using these methods, one can evaluate the impact of gene deletions on the FBA flux rate for a reaction of interest. Genes whose knockdowns significantly impact the production rate of a certain metabolite can be considered for designing nutritional strategies or predicting drug targets. A study proposed in 2011 (Folger et al. 2011) produced a genome-scale metabolic network of cancer metabolism. Using gene deletion and a biomass objective function to simulate growth, the model predicted 52 cytostatic drug targets with 40% of these targets being already used in either approved or experimental treatments. By using FBA-based GEA, common and novel cancer-specific drug targets were also identified in related work (Tobalina et al. 2016). Furthermore, in one reconstruction of *Mycobacterium tuberculosis* similar approaches were used to identify drug targets with a prediction accuracy of 78% (Beste et al. 2007).

It is worth mentioning that the implementation of these tools is based entirely on computational methods. Thus, further experimental validation may be required to investigate accuracy of results.

LP problems can result in multiple optimal solutions for the same objective function. One of the limitations of FBA is that it provides a single optimal solution when an objective function is either maximised or minimised. Some of the alternative optimal solutions can be physiologically meaningful (Reed et al. 2004). Thus, variability of optimal solutions should be assessed (Lee et al. 2000). Flux Variability Analysis (FVA) is a widely used LP approach that calculates minimum and maximum range of fluxes for each reaction flux given a set of constraints and a target objective function. FVA solves two LP problems (a maximisation followed by a minimisation) and can be written as:

$Max/Min\ v_i\ subject\ to$:

$Sv\ =\ 0$

$Z = u^t v \geq \alpha Z_0\ for\ 0 \leq\ \alpha \leq\ 1$

$l_i \leq\ v_i \leq\ u_i\ (LP\ 1.2)$

Where $Z_0 = u^t v_0$ represents the objective function from LP 1.1. The formulation of LP 1.2 was obtained from Gudmundsson et al. (2010). It is important to point out that FVA does not assess all possible optimal solutions (Mahadevan et al. 2003). However, it provides valuable insight into network flexibility and robustness (Gudmundsson et al. 2010).

Having described the principles and concepts behind COBRA approaches, the next part of this literature review will introduce the fibroblast, myofibroblast and skeletal muscle cell.

1.2.5 Biology of the fibroblasts, myofibroblasts and skeletal muscle cells

1.2.5.1 Fibroblasts and myofibroblasts

1.2.5.1.1 Wound healing

The wound healing process has been thoroughly studied since the beginning of human civilisation. For instance, one of the oldest medical records, a clay tablet dated 2200 BC provides a set of procedures for wound care: “washing the wounds, making the plasters and bandaging the wound“ (Shah 2011). Wounds can occur in many tissues types but most of them undergo similar phases in repairing and healing. Typically, normal wound healing is comprised of three stages: inflammatory, proliferative and maturation (also known as remodelling). These stages may, to some extent, overlap one another (Gonzalez 2016).

Wound healing is a dynamic and complex biological process that generates a cascade of cellular and biochemical events. It involves several different cell types and the activation of multiple metabolic pathways. For many years, the research community’s focus was on developing empirical therapeutic strategies rather than understanding the biological mechanisms behind the process of wound healing. More recently, researchers have studied several of the physiological aspects involved in wound healing (Bochaton-Piallat 2016). It is now generally accepted that fibroblastic cells acquire contractile features during normal and pathological tissue repair processes as they differentiate into myofibroblasts playing a central role in wound contraction and scarring. This has opened a new perspective in the understanding of the physiological mechanisms involved in this complex biological process.

1.2.5.1.2 The fibroblast

Fibroblast cells are the most common cells in connective tissue. These cells support the structural framework of tissues by actively producing extracellular matrix (ECM) proteins. Fibroblasts also demonstrate great versatility when compared to other connective-tissue cells. These cells possess 'a remarkable capacity to differentiate into other members of the family' (Alberts et. al 2002). When tissues are damaged, fibroblasts become activated and differentiate into myofibroblasts which facilitate healing by producing ECM proteins and reducing the wound size by contracting local ECM. The proliferation of fibroblasts and their transition to myofibroblasts represent critical events in the process of wound healing and scarring (Li and Wang 2011). It is important to point out that the fibroblast differentiation into myofibroblast is dependent on transforming growth factor (TGF-β), an inflammatory cytokine with multiple cellular roles. TGF-β promotes the fibroblast-myofibroblast differentiation and plays a crucial role in many other regulatory processes including growth, proliferation, apoptosis and ECM production among others (Clark et al. 1998). The role of TGF-β in wound healing, abnormal scarring and fibrosis will be further discussed in sections 1.2.5.1.4 -1.2.5.1.5

1.2.5.1.3 The myofibroblast

Myofibroblasts were initially described in 1971 in the granulation tissue of healing wounds (Gabbiani et al. 1971). They described the structure of these cells ultramicroscopically as a "modified fibroblast" with smooth muscle-like properties. From a morphological point of view, myofibroblasts are larger than activated fibroblasts (Dartt et al. 2010), and present ruffled membranes and a well-developed, active endoplasmic reticulum (Baum et al. 2011). They are characterised by highly organised, contractile bundles of actin microfilaments called stress fibres. The actin microfilaments

terminate in “fibronexus”, an adhesion complex which connects the myofibroblasts’ internal microfilaments to the surrounding extracellular matrix (Eyden 1993, Baum et al. 2011). This creates a contractile mechanism capable of transmitting the force that is generated from stress fibres to the extracellular matrix. Similar to smooth muscle cells, myofibroblasts express α-smooth muscle actin (α-SMA) and incorporate it into intracellular stress fibres. α-SMA is generally accepted as the universal marker for the myofibroblast phenotype (Darby et al. 1990, Tomasek et al. 2002) unless smooth muscle is present. These cells mediate closure of wounds by virtue of their contraction and the formation of collagen-rich scars. However, their sustained presence in organs such as lungs, liver, and kidney beyond wound resolution, and continued matrix synthesis underpins progressive fibrotic disease.

1.2.5.1.4 Myofibroblasts in the process of wound healing

Early in the wound healing process, undifferentiated fibroblasts appear at the wound provisional matrix (Darby et al. 2007) and evolve to proto-myofibroblasts, which are transient cells involved in fibroblast to myofibroblast trans-differentiation (Desmoulière et al. 2005).

Modulation from the fibroblast towards the proto-myofibroblast has been related to the accumulation of stress fibres and mechanical tension, while the switch from the proto-myofibroblast to the differentiated myofibroblast is governed by TGF-β signalling, the presence of specialized ECM proteins (such as the cellular fibronectin (FN) splice variant ED-A) and further mechanical tension generated by the cell’s remodelling activities. Each of these factors alone cannot induce the myofibroblast phenotype and joint action is needed to stimulate the transition from proto-myofibroblast to myofibroblast (Tomasek et al. 2002, Chitturi et al. 2015).

The main difference between proto-myofibroblast and myofibroblast is the α-SMA in the latter (Tomasek et al. 2002) (Fig. 1.7). Myofibroblasts generate contractile forces due to the expression of α-SMA and play a crucial role in wound contraction and closure. The incorporation of α-SMA into stress fibres has been shown to significantly increase the contractile activity of myofibroblasts (Dugina et al. 2001, Hinz 2009). Several studies have shown that the TGF-β induces α-SMA expression (Desmoulière et al. 1993, Tamm et al. 1996).

During the remodelling stage of normal wound healing, the synthesis of ECM is significantly diminished and the synthetic components suffer modifications until granulation tissue disappears completely (Darby et al. 2014).

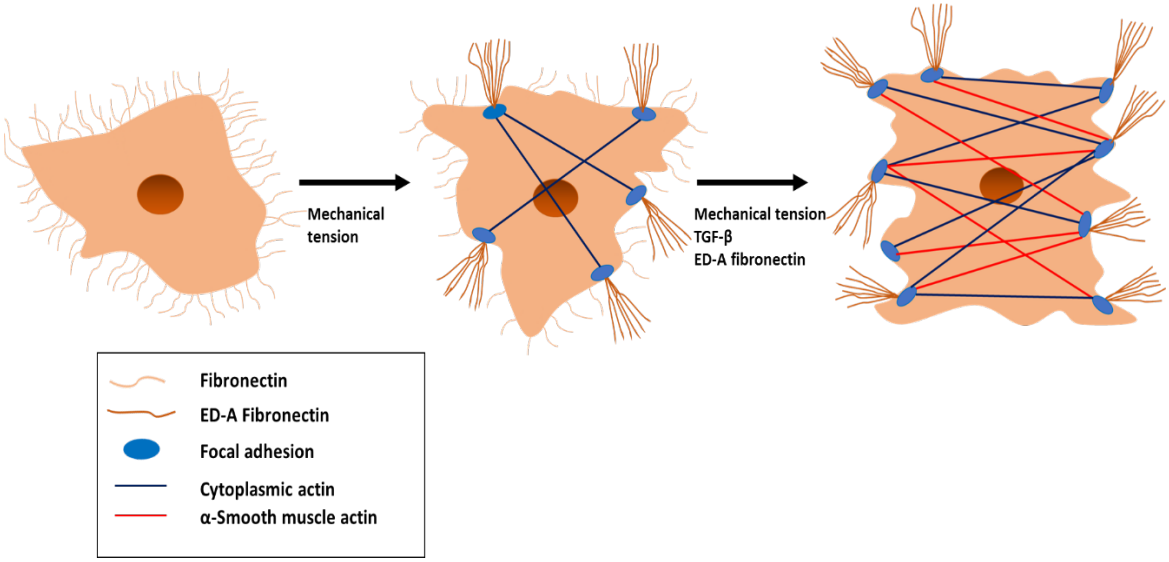


Figure 1.7 Stages of fibroblast to myofibroblast differentiation
Under mechanical stress, fibroblasts acquire stress fibres and evolve to proto-myofibroblasts. The proto-myofibroblast phenotype is characterised by the presence of stress fibres, ECM proteins such as ED-A fibronectin and focal adhesions. In the presence of transforming growth factor β (TGF-β), and further mechanical tension, proto-myofibroblasts can form differentiated myofibroblasts. Myofibroblasts express α-smooth muscle actin (α-SMA). Adapted from Falke, Gholizadeh et al. (2015).

Myofibroblasts typically resolve through a process called apoptosis (Darby et al. 1990, Desmoulière et al. 1995) once the wound proceeds to full epithelialisation, allowing resolution in the absence of an inflammatory response. However, predominantly in

pathological states, myofibroblasts fail to undergo apoptosis and their sustained presence is generally a marker of fibrosis and scarring (Meran et al. 2011, Darby et al. 2014). Fibrotic diseases are responsible for about 45% of all deaths in the western world (Borthwick et al. 2013). Pathological cases resulting in fibrosis are generally characterised by the persistence and progression of highly synthetic myofibroblast cells. This persistence results in abnormal and excessive deposition of ECM which alters the normal structure of the tissue leading to excessive scarring and eventually to a complete loss of organ function. For example, Idiopathic Pulmonary Fibrosis (IPF) is a progressive and irreversible lung disorder that is characterised by a very poor prognosis. In 2012, this fibrotic disease played an etiological role in 0.9% of all deaths in the United Kingdom (British Lung Foundation- www.blf.org.uk) and the median survival time for IPF patients varies from 2 to 3 years after diagnosis (Raghu et al. 2011). This is because the condition is idiopathic and treatment alternatives are very limited.

1.2.5.1.5 Role of myofibroblasts in fibrotic diseases (fibrosis and excessive scarring)

Myofibroblast cells play a key role in abnormal wound healing and drive fibrotic diseases that affect several vital organs such as the kidney, lung, liver, and heart. Persistence of these cells is likely to depend on the local production of TGF-β (Webber et al. 2009). TGF-β not only induces the expression of α-SMA (Desmoulière et al. 1993, Hu, Wu et al. 2003), but also stimulates the accumulation of the stress fibres that are needed to produce mechanical tension. TGF-β also increases the expression of FN (Ignotz et al. 1986, Hocevar et al. 1999) and the focal adhesion complexes (Dugina et al. 2001, Wendt et al. 2009), which are important features of the myofibroblast phenotype (Chitturi et al. 2015). Furthermore, TGF-β alters the ECM production and contributes to the excessive accumulation of ECM proteins including collagens, fibronectin, proteoglycans and elastin by down-regulating matrix degrading enzymes and stimulating their

inhibitors (Leivonen et al. 2013). Early in the wound healing process, the balance between the proteinases that can degrade the matrix and their inhibitors favours ECM production. Later in the healing process, ECM starts a remodelling process and the balance changes, favouring apoptosis and matrix degradation (Lu et al. 2011). In pathological fibrosis, TGF-β stimulates the production of the matrix degrading inhibitors and contributes to the sustained presence of myofibroblast cells at the sites of injury. Therefore, TGF-β directly perpetuates the cycle of ECM deposition and impaired degradation that generally marks the progressive and irreversible remodelling seen in pathological fibrosis. Many studies suggest that future treatment approaches could consider the production of agents that would target and antagonise the activity of TGF-β in fibrotic lesions (Funato et al. 1999, Fernandez et al. 2012). Results have shown that introducing neutralising antibodies to TGF-β *in vitro* considerably reduced the expression of α-SMA in the early stages of wound healing and stopped excessive ECM deposition (Shah et al. 1992, Yokozeki et al. 1997). However, TGF-β is a fundamental biological pathway and treatments targeted at TGF-β might be expected to carry significant toxicity risks (Gueorguieva et al. 2014).

1.2.5.1.6 Metabolic reprogramming in myofibroblasts

Metabolic reprogramming is an early and lasting event that takes place during myofibroblast differentiation (Xie et al. 2015) and plays a crucial role in the contractile function of the cell (Bernard et al. 2015). Myofibroblasts display augmented glycolysis during differentiation and some studies have suggested that TGF-β induces this process (Bernard 2015). The augmented glycolysis phenomenon in myofibroblasts is similar to the Warburg Effect (Warburg 1956) that occurs in cancer cells. It is widely accepted that cancer cells undergo metabolic programming and display augmented glycolysis to sustain proliferation. This phenomenon was extensively studied in cancer cells and

glycolysis is considered a therapeutic target for anticancer drugs (Hamanaka et al. 2012, Ganapathy-Kanniappan et al. 2013). However, the role of metabolic reprogramming in fibrotic diseases has not been fully explored yet. Glycolysis is a ten-step metabolic pathway in which one molecule of glucose is converted into two molecules of pyruvate. Then, under anaerobic conditions, lactate dehydrogenase (LDHA) converts pyruvate into lactate. Three reactions in the glycolytic pathway hexokinase (HK), phosphofructokinase (PFKM), and pyruvate kinase (PKM) enzymes, are irreversible and often considered rate-limiting as they influence the overall glycolytic flux. It was reported that the enzymes catalysing these reactions are up-regulated during *in vitro* myofibroblast differentiation in IPF (Xie et al. 2015). The process of myofibroblast expansion involves two steps (differentiation and proliferation) which are closely related. In the presence of TGF-β, fibroblasts change phenotype and differentiate into myofibroblasts and then TGF-β is also reported to stimulate myofibroblast proliferation. Furthermore, augmented glycolysis is not only essential for myofibroblasts differentiation but might also be required for myofibroblast abnormal proliferation (Srivastava et al. 2018).

Myofibroblasts produce ECM proteins that contribute to tissue repair and wound healing. Among these proteins, collagens are the major structural proteins and considered critical for the wound healing process. In the remodelling stage, collagen type III is gradually replaced by collagen type I (Darby et al. 2014). TGF-β stimulates collagen (type I) production (Chen et al. 1999, Petrov et al. 2002) and diminished collagen type III is associated with myofibroblast differentiation and formation of scars (Volk et al. 2011). One unique aspect of collagen is that glycine, a non-essential AA, accounts for about one third of the AAs in collagen molecule (Barbul 2008).

The *de novo* synthesis of the AA serine and its conversion to glycine (Fig. 1.8) represents an important metabolic pathway that plays an important role in protein, nucleic acids and lipids synthesis. Recently, this pathway was reported to play a crucial role in collagen formation in human lung fibroblasts of IPF (Nigdelioglu et al. 2016). Furthermore, TGF-β was reported not only to stimulate glycolysis but also regulate the expression of genes involved in the biosynthetic pathway of serine and glycine (Nigdelioglu et al. 2016). *De novo* serine synthesis is derived from glucose via glycolysis, and glycolytic intermediate 3-phosphoglycerate (3-PG) is converted to serine through a three-step enzymatic process involving phosphoglycerate dehydrogenase (PHGDH), phosphoserine aminotransferase (PSAT) and phosphoserine phosphatase (PSPH). Then, serine is linked to glycine metabolic pathway synthesis via serine hydroxymethyltransferase (SHMT). The detailed process is as follows: phosphoglycerate (3PG), a glycolytic intermediate, is firstly oxidized (using NAD^+) in a reaction catalysed by phosphoglycerate dehydrogenase (PHGDH), to yield 3-phosphohydroxypyruvate. Then, 3-phosphohydroxypyruvate is transaminated into phosphoserine by phosphoserine aminotransferase (PSAT), using L-glutamate and producing 2-oxoglutamate. Finally, phosphoserine is converted to serine by phosphoserine phosphatase (PSPH) using water and yielding phosphate. Then, glycine is formed from serine in a reversible reaction catalysed by serine hydroxymethyltransferase (SHMT). SHMT enzyme is present in both mitochondria and cytoplasm in the form of two isoenzymes SHMT1 and SHMT2. It was reported that in IPF, serine/glycine biosynthesis is required for collagen production and TGF-β upregulates the enzymes catalysing reactions in this biosynthetic pathway (Nigdelioglu et al. 2016).

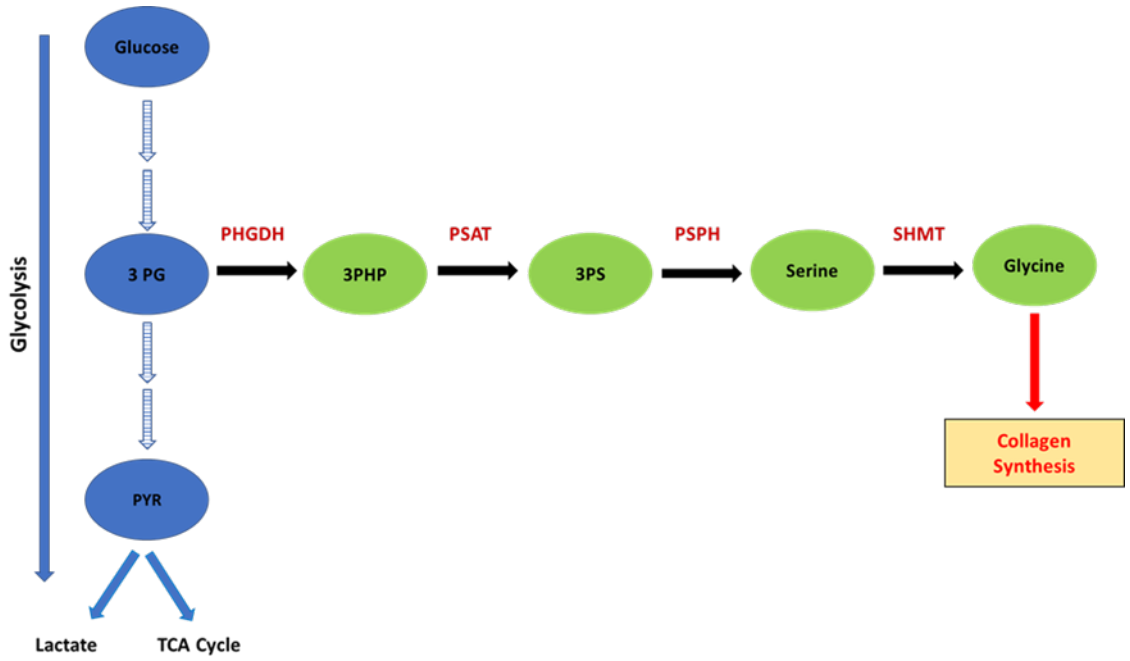


Figure 1.8 Biosynthesis of serine and glycine
Serine is synthesised from the glycolytic intermediate 3-PG. A set of 3 enzymes (PHGDH, PSAT, PSPH) catalyse the reactions constituting this pathway. Then, serine is converted to glycine via serine hydroxymethyltransferase (SHMT). Glycine is one of the main contributors to collagen synthesis, with about one third of collagen molecule being made up of glycine.

A better understanding of the mechanism behind the myofibroblasts' abnormal proliferation and collagen production is of paramount importance as it can provide insight into the causes for its sustained persistence and lead to additional targeted approaches to cure fibrosis and scarring. Thus, the work described here aims to reconstruct, analyse and compare the FB_N and FB_{TGF} models, while focusing on certain biological functions such as proliferation, glycolysis and collagen synthesis.

1.2.5.2 Skeletal Muscle

1.2.5.2.1 Skeletal muscle structure

Muscle can be divided into three distinct types: skeletal, cardiac, and smooth. Each category of muscle tissue has a unique structure and performs different functions. Skeletal muscle is the only muscle tissue under voluntary control through nervous impulses from the cerebral cortex. The functions of skeletal muscle are to enable

movement, provide support and protection to the organs in the body. Skeletal muscle is directly attached to bones by tendons and is composed of long and cylindrical bundles of cells called muscle fibres that contain many nuclei. During development, fibres arise from the fusion of smaller cells (myoblasts) that retain their nuclei. For this reason, skeletal muscle fibre is very often referred to as a 'post-mitotic, multinucleated muscle cell'. Myofibres contain an array of myofibrils that are organised in a region called the "sarcomere" that is the basic functional unit for contraction. Sarcomeres are composed of long protein filaments which allow the muscle cell to contract. These proteins include myosin, actin, and titin which are organised in thick filaments called myofilaments (Jones et al. 2004). Skeletal muscles are sometimes called striated muscles due to their distinctive appearance resulting from regularly arranged protein structures. Under the microscope, they present cross-striated striped pattern due to the distinct and regular arrangement of the myofilaments (Fig. 1.9)

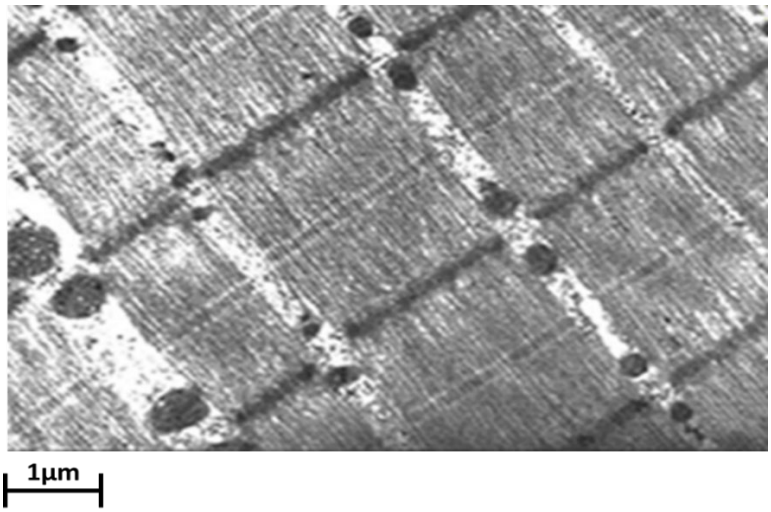


Figure 1.9 An electron micrograph of skeletal muscle showing the interdigitating actin and myosin filaments which make up a sarcomere
Adopted from Sprott et al. (2004) Copyright © 2004, BMJ Publishing Group Ltd.

In each sarcomere, the thick filaments are comprised of myosin, while thin filaments are made up of actin. Thin and thick filaments interdigitate in the sarcomere to facilitate

muscle shortening and lengthening during contraction. The I band is the region of the sarcomere that contains thin actin filaments, while the A band contains thick myosin filaments (Jones et al. 2004). The thin and thick filaments overlap as thin filaments are extended into A band. However, the region in the A band where the thick filaments are not overlapped by the thin ones is called the H zone and contains myosin only (Cooper 2000, Jones et al. 2004). The boundary region of each sacromere is called the Z disc (or Z line) and the middle region in the sarcomere (between two Z lines) forms the M line (Fig. 1.10). Neighbouring sarcomeres share the same Z lines (Cooper 2000). When contraction occurs, the Z lines come closer together and sarcomere shortens. The width of the A band remains constant, I band becomes shorter and H band no longer appears (Cooper 2000, Jones et al. 2004).

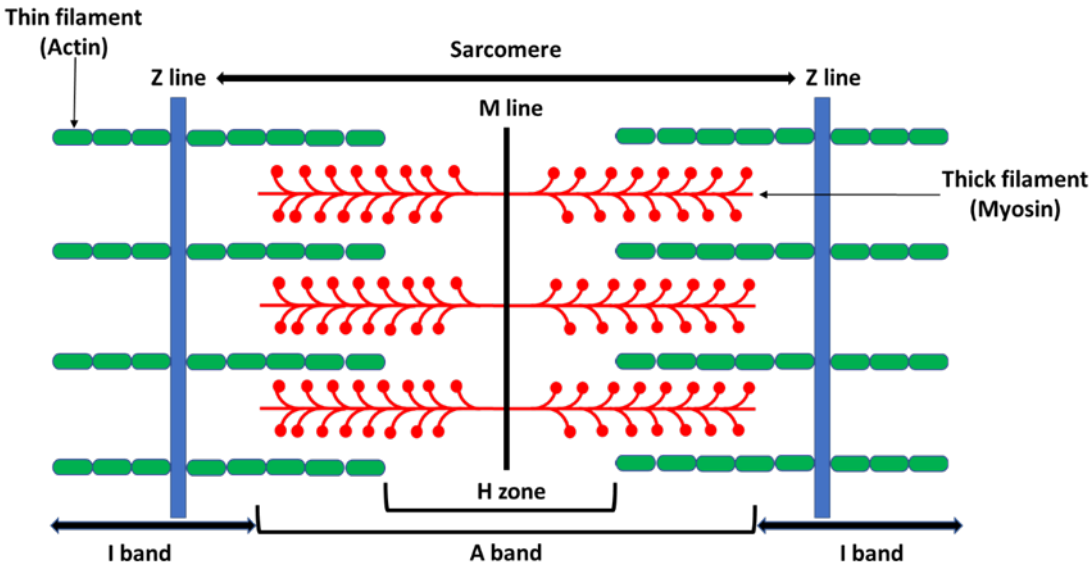


Figure 1.10 Sarcomere structure
*The **I** band contains thin filaments made of actin, while the **A** band contains both actin and myosin filaments. The region in the A band where thick and thin filaments are not overlapped is called the **H** zone and consists of myosin only. Each sarcomere is bordered by two **Z** lines. The **M** line defines the middle of the **A** band and of each sarcomere.*

The sarcomere contains some additional proteins such as titin, nebulin, troponin and tropomyosin. Titin and nebulin are the largest in the sarcomere region and contribute to the sarcomere's structure and stability (Wang 1996, Ottenheijm et al. 2012). Muscle activation and control at the cellular level is through the thin filament complex comprising actin, troponin and tropomyosin. Troponin is itself a complex of three proteins (troponin C, troponin I and troponin T) coupled to tropomyosin. Troponin serves as a calcium regulator for skeletal muscle contraction (Gomes et al. 2002). The binding of calcium to troponin causes the unwinding of tropomyosin which reveals binding sites on actin to which the cross-bridges from the myosin filament can attach and produce force and movement (C. El-Saleh et al. 1986). The ATP is hydrolysed into adenosine diphosphate (ADP) and phosphate, producing energy that is transferred to the heads of myosin filaments. This energy is used by the myosin to pull the actin filaments toward each other from both sides, shortening the sarcomere and producing contraction. When a new ATP molecule binds to myosin, the myosin and actin separate from each other and the cross-bridge detaches (Jones, Round et al. 2004). This process is repeated many times as long as there is sufficient calcium present with ATP, the fuel for contraction being recharged through various metabolic pathways. As there is only about two seconds worth of ATP in a cell, there are a series of processes by which ATP can be generated for the maintenance of contraction which depend on intensity and duration of exercise, balancing the requirements of metabolic power (rate of ATP generation) with capacity (amount of fuel store). These are through substrate level phosphorylation (phosphocreatine (PCr) degradation and glycolysis) and oxidative phosphorylation of carbohydrates and fats during aerobic exercise.

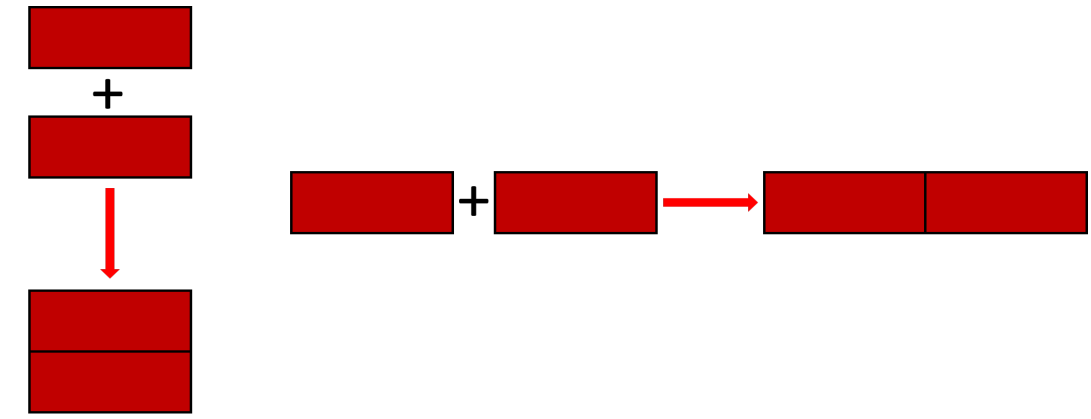
1.2.5.2.2 Skeletal muscle regeneration and growth

Skeletal muscle cell is a post-mitotic cell which means that the repair and regeneration of the skeletal muscle tissue is dependent on the stem-like satellite cells which are located between the basal lamina and plasma membrane of muscle fibres. Whenever mechanical strain occurs, these cells become activated and facilitate repair and regeneration of the muscle tissue. It was reported that satellite cells play a crucial role not only in the muscle regeneration and repair but also in the muscle growth (Phillips 2014, Bazgir 2017). Since muscle fibres are unable to produce nuclei, satellite cells participate in muscle growth by donating myonuclei to existing muscle fibres (Blaauw et al. 2014).

The number of muscle cells is mainly defined in the pre-natal and early postnatal period after which muscle growth occurs by an increase in the size of existing fibres - their diameter or cross-sectional area, a process known as “hypertrophy”. The opposite process is called atrophy and is characterised by a decrease in the muscle mass. Muscle atrophy can occur in health with unloading (such as limb immobilisation, bed rest or microgravity) or in disease such as cachexia (with cancer or HIV) and with ageing (sarcopenia). Muscle growth can also occur when stretch is imposed and the length of a muscle is increased through the addition of sarcomeres in series (Williams et al. 1971). The sarcomere is well-conserved throughout evolution and biology. Although the number of muscle fibres may differ among different species, sarcomeres in humans and animals share the same essential properties, the only difference being the number of sarcomeres that add up in series or in parallel. The architectural design of the muscle has an impact on its functional properties (Lieber et al. 2011). For instance, the way sarcomeres are added up has an effect on the force and velocity of contraction.

Sarcomeres added up in series would mean a faster velocity while those added up in parallel would produce an increased force during contraction (Fig. 1.11).

In the adult, the response to mechanical overload (e.g. resistance exercise) and feeding results in hypertrophy. Here contractile proteins accumulate as the protein synthesis rate exceeds the protein breakdown rate (Tipton et al. 2001). In humans, growth is driven by an increased protein synthesis rather than a reduction in rate of muscle protein breakdown (Carraro et al. 1990). AAs are required to build the proteins and the provision of AAs is reported to be the major stimulant for an increased protein synthesis (Atherton et al. 2012). During hypertrophy there is an increase in actin and myosin filaments within each sarcomere (myofibrillar protein synthesis) and this leads to an increase of the entire contractile unit which is formed by myosin, actin, troponin and tropomyosin. This unit constitutes about 80% of the entire muscle fibre (Wilborn et al. 2004). The most abundant contractile protein is myosin which represents about 25% total muscle protein (Wilborn et al. 2004) while actin represents 10% of the total muscle protein (Cooper 2000). These two proteins are the major components of the contractile unit, playing a crucial role in muscle growth.



A: Sarcomeres in parallel

B: Sarcomeres in series

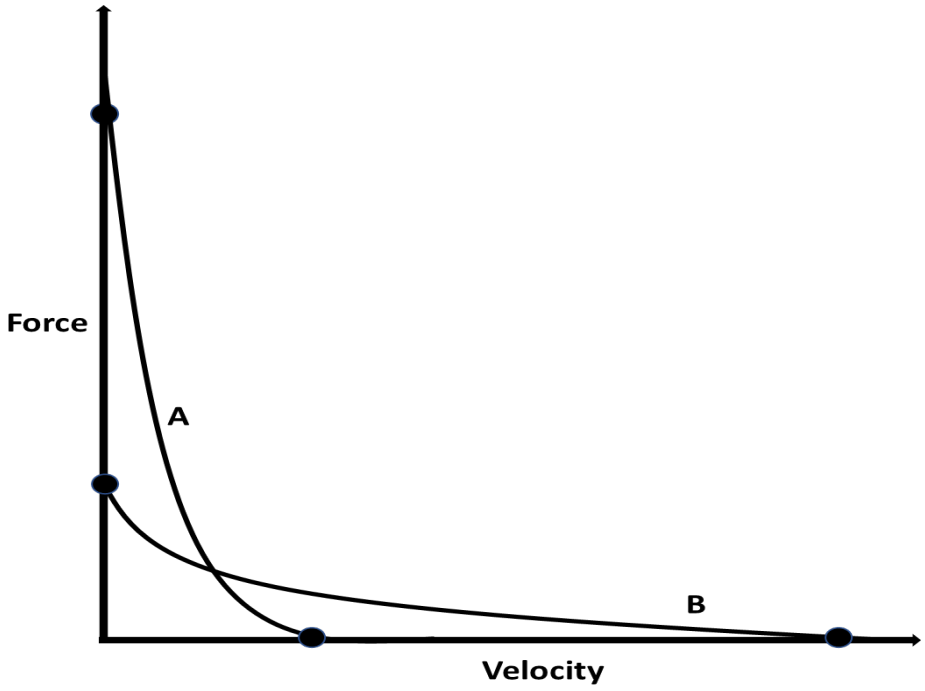


Figure 1.11 Sarcomeres in series vs. sarcomeres in parallel
The force-velocity curves of two hypothetical muscles (A, B) were plotted. A has a high number of sarcomeres in parallel giving high force and low velocity, while B has a low number of sarcomeres in series giving high velocity and low force.

1.2.5.2.3 Hypoxia, protein synthesis and muscle atrophy

Hypoxia can be defined as a deprivation of the amount of oxygen in the whole body or in some tissues. Hypoxia occurs during ascent to high altitude such as climbing a mountain where the oxygen partial pressure (tension) is lowered. Clinically, hypoxia occurs in patients with lung diseases such COPD or lung fibrosis, despite them having

sufficient ambient oxygen. Hence the need for oxygen supplementation in some patients with severe breathing difficulties. Furthermore, tissue hypoxia can also occur in ischaemia when there is insufficient blood flow to maintain tissue viability/normal function.

Hypoxia has been reported to impact body weight (Boyer et al. 1984, Rose et al. 1988), muscle size (Hoppeler et al. 1990) at high altitude and *in vitro* studies suggest that muscle atrophy in hypoxia may be attributed to a reduction in protein turnover (Kraggerud et al. 1995). Although hypoxia has been associated with muscle atrophy at high altitude, its impact on the protein synthesis and AA provision in skeletal muscle has not been fully understood and results remain controversial (Pasiakos et al. 2017). For instance, a lot of additional factors such as diet, amount of sleep and physical activity can impact protein balance and muscle mass during hypoxia exposure at high altitudes. Hypoxia at high altitudes is reported to reduce appetite and perturb the sleep cycle, factors that can impact protein turnover rate. These factors must also be considered when analysing the impact of hypoxia at high altitude (Kucharzewska et al. 2015).

The main mechanisms for ATP synthesis are oxidative phosphorylation and glycolysis. Oxidative phosphorylation, which takes place in mitochondria, is the principal energy provider of the cell and only a small amount of ATP is produced through glycolysis (Fig. 1.12). During hypoxia the amount of mitochondrial ATP is reduced. In consequence, the amount of ATP required for protein synthesis is reduced too and anaerobic glycolysis cannot fully compensate for this loss. It has been reported that in low oxygen environment, skeletal muscle undergoes metabolic reprogramming by favouring anaerobic glycolysis over mitochondrial metabolism (Chicco et al. 2018). This process was initially described by Louis Pasteur in 1857 and is also known as “Pasteur Effect”

(Dixon et al. 1935). This metabolic switch that happens in low oxygen conditions is reported to involve HIF-1 (hypoxia inducible factor 1) which regulates several glycolytic enzymes and inhibits pyruvate dehydrogenase which converts pyruvate to acetyl-CoA (Kim et al. 2006).

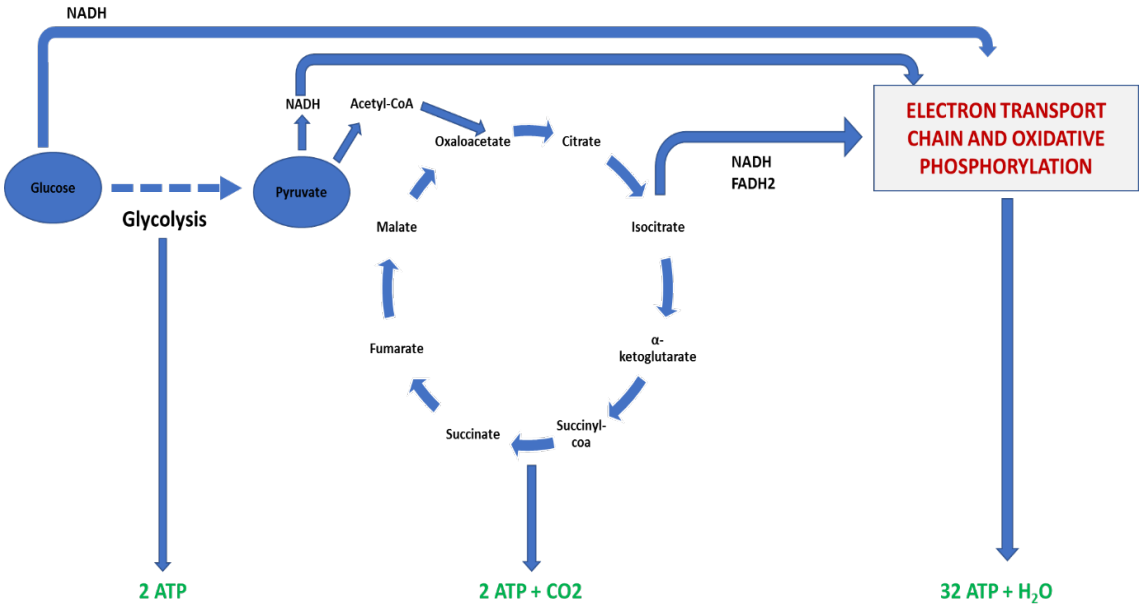


Figure 1.12 A brief overview of the key components of cellular respiration
Glycolysis converts glucose to pyruvate, a process that produces a small amount of ATP and NADH (anaerobic respiration). In aerobic respiration, when oxygen is present, pyruvate formed in glycolysis is converted to acetyl-CoA which enters in the TCA cycle. In the TCA cycle, acetyl-CoA is oxidised and the released electrons and hydrogen atoms combine with NAD^+ and FAD^+ to form NADH and FADH2. The electron carriers (NADH and FADH2) donate electrons to the electron transport chain, which leads to the phosphorylation of ADP, resulting in ATP.

1.2.5.2.4 AA supplementation

The AAs are the building blocks of the proteins. Nine of them (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) are considered essential amino acids (EAA). This means that these AAs cannot be produced by the body and need to be supplied through diet. The other AAs can be synthesised by

the organism. Protein synthesis in the skeletal muscle is dependent on AA availability and each of the AAs, both essential and non-essential, must be provided in the right amounts. Three of the EAAs (isoleucine, leucine and valine) are known as branched-chain amino acids (BCAA) due to their unique structure with side chains containing a central carbon. It was reported that with these AAs, muscles are able to increase protein synthesis rate and stimulate muscle growth (Blomstrand et al. 2006). Furthermore, leucine alone, was reported to play a crucial role in regulating protein synthesis by activating the mTOR (mammalian target of rapamycin) signalling pathway to produce an anabolic response (Garlick 2005, Norton et al. 2006). The anabolic response represents the state in which protein turnover is positive, i.e. protein synthesis rate exceeds the protein degradation rate (Wolfe 2017). Despite what has been reported about leucine, most commercially available protein supplements comprise a mixture of AAs. The hypothesis that some combinations of AAs may augment protein synthesis rate gave rise to a large industry. For instance, a report published by Zion Market Research predicted that global dietary supplements market will be worth 220.3 billion USD by 2022 (Zion Market Research, 2015). However, the results of the scientific reports focusing on AA supplementation remain controversial.

In summary, this literature review has set the background for the modelling approach to be used and described the rationale for the cell types to which the modelling work is to be applied. The specific aims of the work presented in this thesis are thus to:

- Construct two genome-scale metabolic models for the normal fibroblast (control) and TGF-β-stimulated fibroblast (myofibroblast).
- Use constraint-based approaches to study these cells in regard to proliferation, glycolysis and collagen synthesis.

- Construct a genome-scale metabolic model for the skeletal muscle cell.
- Use constraint-based methods to investigate the effects of hypoxia and amino acid supplementation on skeletal muscle metabolism, focusing on protein synthesis rate and muscle atrophy/hypertrophy.

Chapter 2: Materials and methods

This chapter will present the methods used to reconstruct and analyse the CBMs of normal fibroblasts (FB_N), TGF-β-stimulated fibroblasts (FB_{TGF}) or myofibroblasts and skeletal muscle cell. In addition to a description of the computational methods, this chapter will also consider RNA-sequencing (RNA-seq) data sources used in the reconstruction process and the experimental procedures that were used to derive these data.

2.1 Experimental methods

2.1.1 Transcriptomics data

Transcriptomics is the study of the transcriptome, or the set of RNA molecules that are produced by a genome in a cell or tissue. RNA sequencing (RNA-seq) and microarrays are the most common choices for studying the transcriptome in an organism. RNA-seq analysis, also known as whole transcriptome shotgun sequencing (WTSS), uses deep sequencing technologies for transcriptome profiling. RNA-seq has a number of advantages over microarrays (Wang et al. 2009). Unlike arrays, RNA-seq does not depend on the availability of probes that cover already known genes and can detect novel and rare transcripts, as well as splice variants. That is why RNA-seq is more effective in discovery-based studies (Wang et al. 2009). RNA-seq also avoids the hybridization-based issues (background noise, saturation etc) which are usually associated with microarrays. Thus, RNA-seq is more efficient in detecting genes with low expression compared to microarrays (Zhao et al. 2014).

2.1.2 RNA sequencing

RNA-seq was performed by the Centre for Inflammation and Tissue Repair, University College of London (UCL). To obtain these data, confluent monolayers of human primary

lung fibroblasts from healthy volunteers were stimulated (control) with 1ng/ml of TGF-β1 (porcine origin, R&D Systems, USA) for 24 hours using standard Dulbecco's Modified Eagle Medium (DMEM). Total RNA extraction was performed using miRCURY RNA Isolation Kit (Exiqon, Denmark). Poly(A)-tailed RNA enrichment and library preparation were performed using the KAPA Stranded mRNA-Seq Kit with KAPA mRNA Capture Beads (KAPA Biosystems, Wilmington, MA, USA). Data were stranded and paired-end sequenced to a depth of 15M using NextSeq (Illumina, San Diego, CA, USA). Reads were pre-processed and aligned to the reference genome (*Homo sapiens* UCSC hg19) using STAR aligner (Selvarajah et al. 2019).

2.2 Computational methods

2.2.1 Programs and software packages

R (R Core Team 2013) is a free and open source programming language that is used for statistical computing, graphics, data analysis and data visualisation. R was used to process and manipulate gene expression data. BiomaRt (Durinck et al. 2009) is a Bioconductor (www.bioconductor.org) package that provides an R interface to several public biological databases such as Ensembl (Hunt et al. 2018), COSMIC (Tate et al. 2018), UniProt (Consortium 2018), HGNC (Yates et al. 2017), Gramene (Tello-Ruiz et al. 2018) or WormBase (Harris et al. 2019). In this thesis, BiomaRt was used to convert Ensembl gene IDs to HUGO gene IDs. Custom Python scripts were used to perform data manipulation tasks such as counting the occurrence of an object in list or joining data sets. DeSeq2 (Love et al. 2014), another Bioconductor package, was used for the normalisation and log2 transformation of the gene expression values and to perform differential gene expression analysis. MATLAB (The MathWorks, Natick, MA, USA) is a high-level programming language that can be used for algorithm implementation and

numerical computation. It is used for solving linear algebra problems and has a rich set of matrix and array operations. MATLAB was used as the principal programming language to build CBMs and run simulations. COBRA Toolbox (Becker et al. 2007) is a software package which contains several comprehensive MATLAB scripts useful for constraint-based modelling. These scripts can be used to build and analyse metabolic networks using constraint-based methods. In this thesis, the COBRA Toolbox was used to construct, validate and simulate three genome-scale models for the FB_N, FB_{TGF} and skeletal muscle cell. CPLEX (IBM ILOG CPLEX) is a high-performance linear programming solver that can deal with Linear Programming (LP), Mixed Integer Programming (MIP) and Quadratic Programming (QP) problems. CPLEX was used to solve LP problems that are part of the FASTCORE family of algorithms and for the optimisation of several objective functions of the model.

2.2.2 Differential expression analysis

Differential expression analysis (DEA) performs statistical analysis to verify whether a gene's expression varies between two or more conditions based on the normalised expression data from a small number of replicates. In this framework, DEA was performed using the Deseq 2 package in R to analyse the expression of the genes in the "control" and "TGF-β-stimulated" conditions. Four replicates were used for each condition.

Deseq 2 models the raw count data by first estimating the size factors to account for the library depth. Next, the gene-wise dispersions are estimated, and a curve is fitted to these estimates. To produce more accurate results, the gene-wise dispersion estimates are corrected using the average expression strength over all samples. It is important to point out that Deseq 2 assumes that genes of similar average expression levels have similar dispersion. Next, Deseq 2 fits a negative binomial general linear model (GLM)

that produces the final $\log_2$ fold change. Finally, the fold changes are shrunk according to the mean and dispersion of a gene. Thus, genes with very low counts are shrunk towards zero. Finally, the hypothesis testing is performed using the Wald-test with the null hypothesis stating that there is no significant difference between the average expression levels of the two conditions for a certain gene. Thus, the *p* and *p-adjusted* values are obtained for every gene.

RNA-seq data was loaded in R and the Ensembl gene ids were converted to the corresponding HUGO gene symbols. After the successful implementation of the Deseq2 protocol (Love et al. 2014), a table containing gene names (as row names) together with the corresponding p-value, p-adjusted value and log2 fold-change value was obtained.

Firstly, genes were sorted in ascending order by the p-adjusted value to identify those genes with the most significant fold-change. Then, a volcano plot was used to analyse the results from the DEA. A volcano plot is a type of scatter plot that displays "fold-change" versus "significance". In this way, one can plot the genes with the largest fold changes which are also most statistically significant. In a volcano plot, the X-axis shows the log2 fold-change value, while the Y-axis represents negative $\log_{10}$ of the p-value. Genes of particular interest for the analysis of the models were displayed on the volcano plot to investigate whether they were up-regulated or down-regulated under TGF-β stimulation. The results of DEA are described in section 3.3.5 from Chapter 3.

2.2.3 Processing RNA-seq data for model generation

Data sets generated from RNA-seq analysis were loaded into MATLAB. Entrez genes were mapped to the equivalent HUGO Gene ID used in Recon2.2. MATLAB and HGNC data set (HUGO Gene Nomenclature Committee at the European Bioinformatics Institute) were used in this step. Recon 2.2 (Swainston et al. 2016) was used as a basis for generating a genome-scale model. Recon 2.2 consists of 1673 genes, 7785 reactions

and 5324 metabolites. The GPR associations in Recon 2.2 were used to identify active (core) reactions in the required biological context and FASTCORMICS (Pacheco et al. 2015), an adaptation of FASTCORE algorithm (Vlassis et al. 2014), was used to generate context-specific models.

2.2.4 FASTCORE

FASTCORE is a family of algorithms that can be used to generate context-specific models. The FASTCORE family is comprised of FASTCORE (Vlassis et al. 2014) and three other algorithms: FASTCORMICS (Pacheco et al. 2015), FASTCC (Pacheco et al. 2015) and fastgapfill (Thiele et al. 2014). FASTCORE considers a *core* set of reactions that have strong evidence to be active in a certain biological context and then adds a minimum set of reactions from the global network to the initial core set in order to create a context-specific model. For instance, this problem can be summarized as follows: Given a global network with $\boldsymbol{A}$ reactions and stoichiometric matrix $\boldsymbol{S_A}$ and a set $\boldsymbol{B} \subseteq \boldsymbol{A}$, find the smallest possible set $\boldsymbol{C} \subseteq \boldsymbol{A}$ such that $\boldsymbol{B} \subseteq \boldsymbol{C}$ and all reactions in the subnetwork ($\boldsymbol{S_C}$, $\boldsymbol{C}$) can carry a non-zero flux in at least one flux distribution (network consistency)**.** Thus, the algorithm considers a core set of reactions $\boldsymbol{J} \subseteq \boldsymbol{B}$ with high confidence of being active and a penalty set $\boldsymbol{P} = (\boldsymbol{A} \backslash \boldsymbol{B}) \backslash \boldsymbol{C}$ formed of non-core reactions. At every step, the algorithm solves two LP problems: one for the maximisation of the flux through as many core reactions as possible (LP 2.1), and another one for the minimisation of the flux through reactions which are not included in the core (LP 2.2).

It is worth mentioning that FASTCORE aims to include in the reference network all reactions that belong to the core set (Vlassis et al. 2014). Thus, FASTCC checks the consistency of the network and identifies flux-inconsistent reactions. In this framework, network consistency is valid when every reaction has a flux of at least $\boldsymbol{\varepsilon}$. The value for $\boldsymbol{\varepsilon}$

can be varied and it often depends on the magnitude of the uptake/secretion rates in the model. Investigating the consistency of a network is a commonly used method the identification of potential flaws and deficiencies in a metabolic network (Acuña et al. 2009). FASTCC is a variant of FASTCORE which considers all reactions in the network to be part of the core set and has no penalty set. If the network is not consistent, FASTCC provides a list of flux-inconsistent (blocked) reactions. FASTCC is comprised of the following major steps:

- Step 1: Consider all reactions in the core by default and maximises flux through as many core reactions as possible.
- Step 2: Take all reactions in the core set that carry a flux greater than or equal to a threshold $\boldsymbol{\varepsilon}$ and force them to carry flux, while minimising the overall flux in the network.
- Step 3: Reactions that could carry flux in step 2 are added to the model and removed from the core. Irreversible reactions that couldn't carry flux are identified and considered blocked reactions.
- Step 4: For each remaining reversible core reaction: Reverse the directionality of the reaction (flip) and test whether it can carry flux. If the reaction is carrying flux, remove it from the core set and add it to the model. Reactions that are still unable to carry flux are considered blocked reactions.

These steps can be summarised by the following LP problem:

$$max \sum_{i \in J}^{\infty} z_i$$

$$Sv = 0$$

$so\ that\ z_i \in [0, \varepsilon]\ for\ any\ i\ \in J$ and a small positive threshold ε

$v_i \geq z_i\ for\ any\ i\ \in J$

$l_i \leq v_i \leq u_i$ $(LP\ 2.1)$

Where $\boldsymbol{J}$ is a subset of B (core set) and l_i, u_i represent the lower and upper flux bounds. This LP problem is sufficient to assess the irreversible reactions and identify the blocked ones. However, it works only for positive fluxes. For negative fluxes, an iterative approach is taken. Reversible reactions are assessed one by one using LP 2.1, in both directions and this is achieved by flipping signs of elements in the columns of the stoichiometric matrix. It is important to point out that LP 2.1 is the same as the first LP problem solved by FASTCORE at every step. The second LP problem (LP 2.2) aims to minimise the L1 norm of vector v_i ($\sum_{i \in P} |v_i|$) in the penalty set $P = (A \backslash B) \backslash C$ subject to a minimum flux for the active reactions in new subset $K \subseteq J$. K is obtained after a single run of LP 2.1. The mathematical formulations of LP 2.1 and LP 2.2 were adapted from Vlassis et al. (2014).

$$min \sum_{i \in P}^{\infty} z_i$$

$$S_A v = 0$$

$$so\ that\ v_i \in [-z_i, z_i]\ for\ any\ \text{i} \in P$$

$$v_i \geq \varepsilon\ for\ any\ for\ any\ i\ \in K$$

$$z_i > 0$$

$$l_i \leq v_i \leq u_i\ (LP\ 2.2)$$

Fastgapfill (Thiele et al. 2014) is an algorithm that uses FASTCORE concepts and aims to identify and fill potential gaps in the genome-scale metabolic model. Fastgapfill uses the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2000) as a reference network in this process. Firstly, fastgapfill identifies blocked reactions and dead-end metabolites based on the reference network, and then fills the gaps with metabolic, exchange and transport reactions that will connect dead-end metabolites

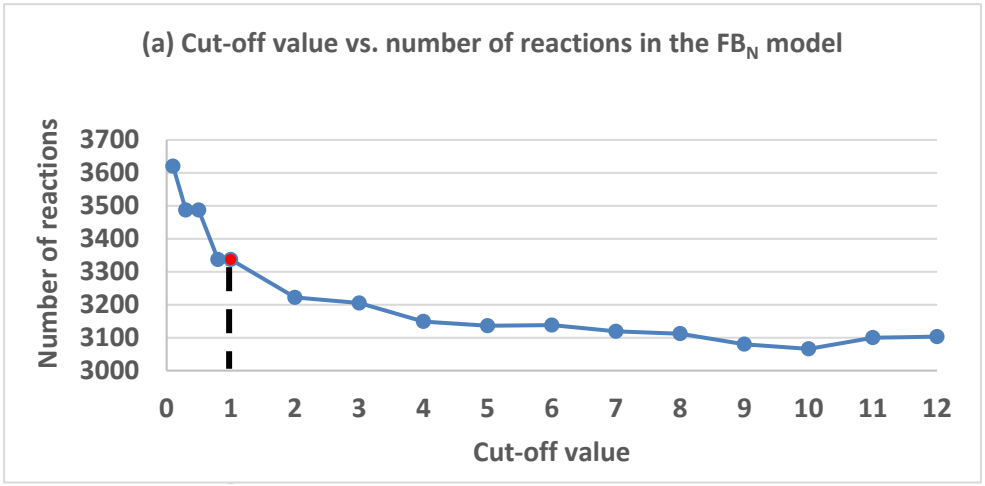
and restore network consistency. It is important to point out that one can prioritise the type of reaction that should be added. In general, the desired order is metabolic over exchange and transport reactions.

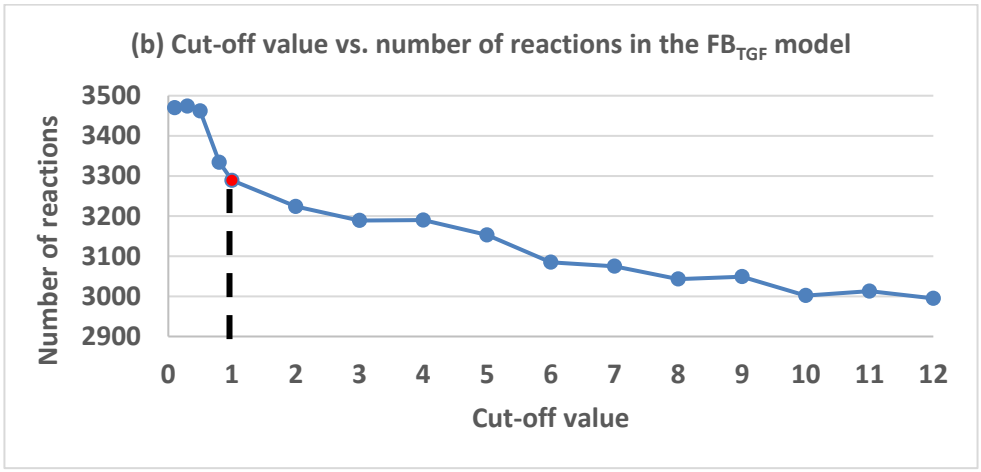
In order to map gene expression data to GPRs in the global reconstruction network, fixed thresholds are commonly imposed to distinguish between active and inactive genes. FASTCORMICS (Pacheco et al. 2015) is an adaptation of FASTCORE that uses Barcode (McCall et al. 2011) as a discreditation tool. Expression data is converted into z-scores (the number of standard deviations) using Barcode and then compared to a recommended threshold of 5z. Also, FASTCORMICS introduces an unexpressed threshold of 0 (standard deviation). This prevents the inclusion of reactions regulated by unexpressed genes. Those genes with 90% of arrays greater than 5z are assigned 1, while those with 90% arrays lower than unexpressed threshold (0) are assigned -1. Genes with a score of 1 are included in the core set while those with a score -1 are removed from the model. Then, a slightly different version of FASTCORE is run to maximise flux through core reactions and penalise the non-core ones. Unlike traditional FASTCORE, this modified version also allows for a set of not-penalised reactions. By heuristics, transport reactions are removed from the core, included in the non-penalty set and favoured over non-core ones. The reason for this is because transport reactions are usually regulated by promiscuous genes that control many other reactions. Furthermore, one can add reactions of interest in the core set and force them to carry flux. This set of “forced” core reactions is known as essential reactions set. This set is manually defined by the user and is comprised of those reactions which were removed by FASTCORE but were reported in the literature to be present, or reactions that were recently added to the model (e.g. biomass reaction, collagen synthesis, etc.), which one wants to make sure are picked by FASTCORE in the final model output. After this is done,

FASTCORE is run once with all essential reactions (the set defined by the user, not the core set determined by the algorithm). Then, the reactions in the "to remove" set are removed, and FASTCC is run on the remaining model. If any reaction from the essential reactions set cannot carry flux anymore, all reactions from the previous solution above are re-added. All core reactions, which are no longer in the consistent model are then removed from the core. The assumption is that reactions from the essential reactions set are prioritised over the active ones. Finally, in the last step FASTCORE is run with the remaining core reactions, while not punishing flux through the non-penalty set.

2.2.5 Initial model generation

Barcode was initially designed for microarray data. Thus, FASTCORMICS was adapted with a different pre-processing step to determine the active and inactive reactions in the context-specific model. For the FB_N and FB_{TGF} data, gene expression values (counts) were normalised using the R package DeSeq2, then log2 transformed with an offset of 1, and genes with count > 1 were called present and assigned 1, while all other genes were called absent and assigned -1. Essentially, corresponding reactions to absent genes (-1) were evaluated as inactive and then placed in the "to remove" set. Similarly, the active genes (1), were used to determine active reactions and placed into the core set. For the skeletal muscle data, the same discreditation process was applied for the transcripts per million (TPM). A cut-off value of 1 was chosen based on sensitivity analysis for all three data sets: FB_N, FB_{TGF} and skeletal muscle (Fig. 2.1). Sensitivity analysis is a mathematical method to assess the robustness of a model or system with respect to the changes of certain parameters.


(a) Cut-off value vs. number of reactions in the FB$_N$ model
Number of reactions
3700
3600
3500
3400
3300
3200
3100
3000
0 1 2 3 4 5 6 7 8 9 10 11 12
Cut-off value


(b) Cut-off value vs. number of reactions in the FB$_{TGF}$ model
Number of reactions
3500
3400
3300
3200
3100
3000
2900
0 1 2 3 4 5 6 7 8 9 10 11 12
Cut-off value

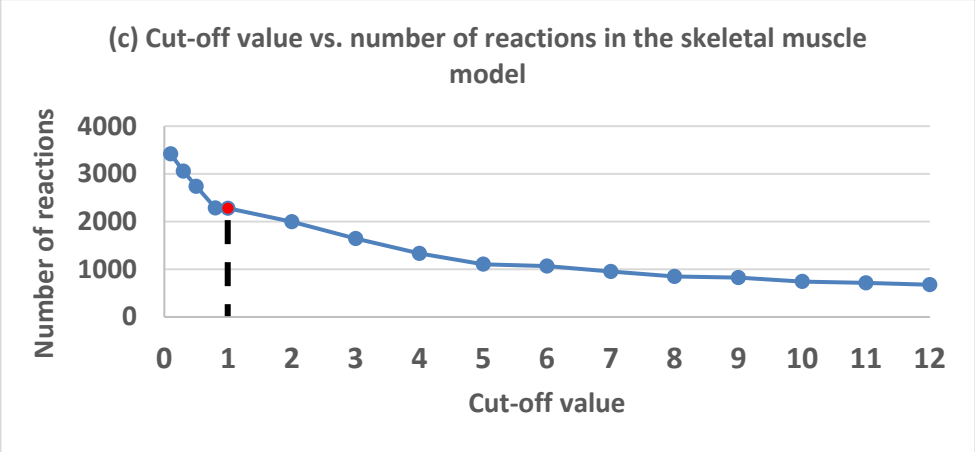


Figure 2.1 (a) (b) (c) The relationship between cut-off value and the number of reactions in the model
The cut-off value was varied between 0.1 and 12 and for each cut-off value (either counts or transcripts per million), the number of reactions was determined. After 1 there is a relatively steady decline in the number of reactions (1 becomes an inflexion point). Thus, a cut-off value of 1 was chosen and all genes with expression below 1 (between 0 and 1) were removed from the model.

2.2.6 Manual refinement

Despite the improvement in automated reconstruction methods it is widely accepted that manual refinement is still needed (Thiele et al. 2010). Thus, the model was manually inspected, curated and refined with the aid of literature data. FASTCORE is a very "restrictive" algorithm and only allows reactions which are necessary for the use of the core to stay in the model. This is in some ways a drawback of FASTCORE as this algorithm tries to generate minimal networks so if some exchanges are not part of the core set, they are likely to be removed. One idea of FASTCORE is to not add unnecessary reactions and rather generate a small model of highly active metabolism. Thus, missing reactions (mainly exchange ones) were added either manually or by using the essential reaction set that forces missing reactions to carry flux. Next, FASTCC was run to identify flux inconsistent reactions. Finally, flux inconsistent reactions were removed from the model (Fig. 2.2). Furthermore, several constraints were imposed on exchange reactions to improve model contextualisation. These constraints were estimated using literature

values and are further detailed in sections 3.2.1 and 4.2.2. Tables 6.2.1-6.2.3 in the Appendices section contain the list of constraints for FB_{TGF}, FB_N and skeletal muscle cell models.

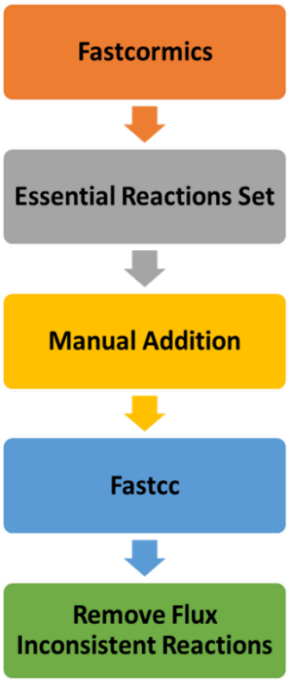


Figure 2.2 Model generation process and manual refinement
The curation process starts with the definition of the essential reaction set which forces certain reaction to become part of the "core" set. Next, the curation process continues with the manual addition of certain reactions. Finally, FASTCC is run for the last time to remove the flux inconsistent reactions which might be generated by the manual addition of certain reactions.

2.2.7 Model analysis

2.2.7.1 Flux balance analysis

FBA (Orth et al. 2010) was used to find flux distributions that optimise certain objective functions in the model. However, FBA provides a single solution for the optimisation problem and alternate possible solutions should be assessed.

2.2.7.2 Flux variability analysis

FVA (Mahadevan et al. 2003) was employed to calculate the maximum and minimum flux range in the steady-state feasible solution space. FVA allows for the same objective

function and constraints as for FBA optimisation. Essential reactions were identified using flux spans. Flux spans were derived using the absolute difference between maximum and minimum fluxes generated by FVA. Then relative flux spans were calculated and expressed as ratios. Flux span ratios were defined as the absolute value of the difference between maximum and minimum flux generated by FVA, divided by the absolute value of the FBA flux. These ratios were sorted in ascending order and those reactions with the lowest flux span ratio were considered essential to the objective function (reaction). It is important to point out that this method is not able to put reactions into categories (essential or not essential). However, the ascending sort order of flux span ratios can suggest which reactions are most "sensitive" to a certain objective function. Not surprisingly, exchange and transport reactions for AAs, water, oxygen, etc. feature heavily in any list of essential metabolic reactions. Those reactions were excluded from the list as their essentiality to a certain objective function is expected.

2.2.7.3 Gene essentiality analysis

FBA-based GEA (Edwards et al. 2000) was used to identify essential genes for the required objective functions. Essential genes were defined as those genes the deletion of which will significantly impact the objective function. In this framework, essential genes were defined as those genes whose deletion would completely disable an objective function (generate a zero FBA flux). GEA can identify single essential genes (single gene deletion) or pairs of essential genes (double gene deletion). Critical genes were defined as those genes whose deletion will considerably reduce the flux for the objective function.

2.2.7.4 Robustness analysis

Robustness analysis (Edwards et al. 2000) was used to investigate and visualise how changes in the uptake rates of a certain reaction impact the objective function.

2.2.8 Objective functions

Optimisation requires an objective function. Three objective functions, simulating reactions which are central to the modelling objectives, were constructed. They are biomass production, collagen synthesis and myofibrillar protein synthesis (production of Acto-Myosin complexes).

2.2.8.1 Biomass reaction

During proliferation, cells modify their metabolism to meet the requirements of this biological function and support macromolecule and precursor synthesis. Biomass rates are usually tightly coupled to proliferation rates (Vaidyanathan et al. 2006) and metabolic processes behind biomass objective function are closely related to cell proliferation (Fajas et al. 2018). Optimising a biomass reaction can provide an insight into the cells' metabolic circuits during proliferation.

To build a biomass reaction for the two models, the macromolecular content of the fibroblasts was derived from the relevant literature (weight fraction of protein, lipids, sugar, RNA and DNA): protein (Galacci and Higuchi, 1966), lipids (Macarak et al. 1979, Harzer et al. 2001), sugar (Umapathysivam et al. 2005), RNA (Miltenyi Biotec GmbH, 2009) and DNA (Bordbar et al. 2010). The macromolecular weight percent contribution of each constituent was calculated using well-accepted methods (Bordbar et al. 2010). Biomass calculations in Bordbar et al. (2010) were used as a template for our calculations. Then, each macromolecule was broken down into smaller building blocks or components that are required to synthesise the macromolecule in terms of molar fractions. Thus, lipids were broken down in two fractions of the total lipids: neutral lipids and phospholipids. Neutral lipids were broken down into individual fatty acids that are present in the chemical processes from the cell metabolic network. The molecular

weight for every fatty acid was calculated and incorporated in the biomass reaction. Sterols, triglycerides and diglycerides were also included in the biomass composition. Major phospholipids (phosphatidylcholine, phosphatidylethanolamine, sphingomyelin, phosphatidylserine/phosphatidylinositol, phosphatidylglycerol, lyso(bis)phosphatidic acid) were considered and their contribution to the biomass reaction was calculated. Also, proteins were broken down into AAs and considered in the biomass formulation. Glycogen was considered the main source of sugars. ATP maintenance requirements, proportion of DNA and RNA monomers were also calculated and incorporated in the biomass reaction. All calculations for the biomass reaction are further detailed in Tables 6.1.1-6.1.7 in the Appendices. The literature sources are included in the table. It is important to note that when no human fibroblast data was available, rat/chicken fibroblast data was used to fill the gaps. Human reconstruction (Recon 1) was previously used to generate a genome-wide model for mouse metabolism with excellent predictive capabilities. Furthermore, it was reported that human physiology is more related to farm animals than to rodents (Hamernik 2019) and chicken provides a suitable model for the study of human metabolism or disease (Schock et al. 2016, Hamernik 2019). Finally, human macrophage data (Bordbar et al. 2010) was considered when no fibroblast parameters of any kind were found.

2.2.8.2 Collagen synthesis

A collagen production reaction was added to Recon 2.2. The AA sequence for collagen (type I) was obtained from UniProt (www.uniprot.org). Every letter in the sequence accounts for an AA. The occurrence of every letter was counted using the Python programming language and the total counts were incorporated in the collagen reaction as stoichiometric coefficients for the corresponding AAs. Energy requirements were adapted from a biochemistry problem proposed in (Berg et al. 2002)). In this problem it

was assumed that the hydrolysis of inorganic phosphate (PP_i) is equivalent to the hydrolysis of ATP, i.e. $ATP \rightarrow AMP + PP_i$ and $PP_i \rightarrow P_i$ provides the same energy. Thus $ATP \rightarrow AMP + 2P_i$ gives double the energy as $ATP \rightarrow AMP + PP_i$. The same principle applies for $GTP \rightarrow GMP + 2P_i$. Thus, to combine n AAs, the energy equivalent of $2n\ x\ (ATP \rightarrow AMP + PP_i)$ hydrolyses is needed. This is equivalent to $nATP \rightarrow nAMP + 2nP_i$.

Furthermore, $1\ GTP \rightarrow GMP + 2P_i$ is required for initiation along with the energy from $(2n-2)\ x\ (GTP \rightarrow GMP + PP_i)$ for the peptide bonds. This is equivalent to:

$(n-1)GTP \rightarrow (n-1)GMP + 2(n-1)P_i$. In total, $nGTP \rightarrow nGMP + 2nP_i$ are required to combine n AAs.

To sum up, $nATP \rightarrow nAMP + 2nP_i$ and $nGTP \rightarrow nGMP + 2nP_i$ are needed to combine n AAs. Thus, for n AAs, the general equation is:

$$\boldsymbol{nAAs + nATP + nGTP \rightarrow nAMP + nGMP + 4nP_i + Collagen}.$$

where n represents the sum of the stoichiometric coefficients for every AA in the reaction. For the collagen synthesis reaction $n = 1446$

2.2.8.3 Acto-Myosin Complex in skeletal muscle

Actin and myosin are the major contractile proteins in the skeletal muscle and for a muscle to grow, these protein filaments must both increase in number and proportion and be assembled in a manner which retains the structural make-up of the sarcomere. The number of myofibrils in a muscle increases, which in turn leads to an increase in the number of these contractile proteins. In humans, contractile protein synthesis is the key

process involved in the muscle mass regulation. Thus, it is important that any model constructed should focus on the synthetic arm of this process.

In order to do this, a system of reactions that simulated the contractile protein synthesis during muscle growth was constructed. Firstly, the AA sequences for actin (alpha skeletal muscle) and myosin (type I) were downloaded from UniProt (www.uniprot.org). Then, similarly to that undertaken and described with collagen in section 2.2.7.2, letters in the AA sequence were counted using Python and the occurrence of every letter was incorporated as a stoichiometric coefficient for the corresponding AA. Energy demands were also considered in the reaction (as described in section 2.2.7.2). The general form of the reactions for both actin and myosin are:

$\boldsymbol{nAAs + nATP + nGTP \rightarrow nAMP + nGMP + 4nP_i + Actin}$ and

$\boldsymbol{nAAs + nATP + nGTP \rightarrow nAMP + nGMP + 4nP_i + Myosin}$

where n represents the sum of the stoichiometric coefficients for every AA in the reaction. For actin building reaction $n = 377$, while for myosin building reaction $n = 1939$.

Then, the stoichiometric (molar) ratio for actin and myosin was considered to be 6:1 in the contractile complex (Murakami et al. 1985). Finally, a demand reaction for the production of Acto-Myosin complex was considered. This reaction served as the main objective function for the model simulations.

Chapter 3: Reconstruction and analysis of fibroblast and TGF-β-stimulated fibroblast cells using constraint-based methods

3.1 Introduction

Wound healing is a complex and dynamic process that involves multiple cell types. Fibroblasts play a key role in the process of wound healing. When injury occurs, these cells migrate to the site of the injury and differentiate into myofibroblasts which contract the wound by virtue of their contractile properties. Myofibroblasts express α-SMA, which is considered the primary marker for the myofibroblast phenotype and TGF-β has been shown to regulate the expression of this actin (Desmoulière et al. 1993). Following wound repair, myofibroblasts undergo apoptosis. However, pathological cases are characterised by myofibroblasts' sustained presence, abnormal proliferation of these cells and excessive collagen production. In fibrotic diseases, TGF-β is reported to stimulate these biological functions. (Verrecchia et al. 2007).

A metabolic model of the myofibroblast cell would be important for identifying metabolic pathways of chemical reactions in myofibroblast cells that might serve as therapeutic strategies to inhibit both proliferation and collagen deposition. Here two genome-scale metabolic models for FB_N and FB_{TGF} are presented. FB_{TGF} is a commonly used *in vitro* model of the myofibroblast cell (Desmoulière et al. 1993). The two models were analysed separately and then compared using constraint-based tools such as FBA, FVA and GEA. Many of the predictions generated by computational simulations were independently validated *in vivo*. These models could illustrate a fundamental computational counterpart for interpreting transcriptomics data characterizing myofibroblast metabolic alterations in IPF. Identifying and understanding IPF hallmarks

is crucial for designing new therapeutic targets that will limit myofibroblast proliferation and collagen production.

3.2 Methods

3.2.1 Manual refinement

A manual refinement of the models was performed as described in section 2.2.5. Due to the "restrictive" nature of FASTCORE algorithm, certain reactions can be excluded from the model. In this scenario, missing reactions that were not picked up by FASTCORE can be "forced" into the model. Firstly, the models were constructed without any essential reactions set. Then, after inspection, reactions such as essential exchanges were added in the essential reactions set and forced to carry flux. Furthermore, new reactions such as collagen synthesis or biomass reactions were also included in the set. In this framework, the essential reaction set was comprised of biomass reaction, collagen synthesis reaction and the exchange reactions for collagen, lactate and water.

Subsequently, constraints (upper and lower bounds) were imposed for glucose, oxygen, and AA uptake and lactate secretion exchange reactions to improve model contextualisation. The majority of non-essential AAs were considered to have no influx (a lower bound of zero) as the model should be able to produce them (internally). An exception was made for glutamine, serine, asparagine and aspartic acid which were reported to be consumed by the cells (Dietmair et al. 2012). Hence these amino acids were made available in the model by imposing a non-zero lower bound. All the constraints for the uptake/secretion rates were derived from literature: oxygen (Bernard et al. 2015), AAs, glucose, lactate (Dietmair et al. 2012). For some rates, data from human embryonic kidney 293 (HEK 293) cells were used (Dietmair et al. 2012). These cells, which although of uncertain origin with respect to parent cell type, have

some morphological features of fibroblasts and may be of that origin (Stepanenko et al. 2015). All rates were converted to mmol gDW^{-1} h^{-1}, which was the chosen unit for fluxes. Where necessary, transformations were performed to obtain this unit. In order to perform some of these transformations, the dry weight of a single myofibroblast (843 picograms) was derived from literature (Lee et al. 1977). As substrate uptake rates for FB_N and FB_{TGF} are significantly different, where possible, individual constraints for FB_{TGF} fibroblast cells were used. In the absence of data, the uptake/secretion bounds in the FB_{TGF} model were estimated using the bounds in the FB_N model. TGF-β was reported to increase AA uptake in lung fibroblasts by approximately two fold (Subramanian et al. 2006). Thus, lower bounds for AA exchanges in the FB_{TGF} model were calculated by doubling the constraints for FB_N. Furthermore, *in vivo* studies also reported a 2.5 increase in lactate production (Zhang et al. 2015) and an increase by approximately two fold in glucose uptake for FB_{TGF} when compared to FB_N. Thus, the lower bound for glucose uptake and the upper bound for lactate efflux in the FB_{TGF} model were derived in the same way as AAs (by doubling the bounds from the FB_N model). The constraints for the remaining reactions in the model were not amended. Tables 6.2.1-6.2.2 in the Appendices section contain a complete list for the imposed constraints in the FB_N and FB_{TGF} models. Lastly, FASTCC was run again to identify flux inconsistent reactions. These reactions were then removed from the model. A threshold of 1×10^{-6} was chosen for the FASTCC workflow.

3.2.2 Model analysis

A biomass objective function was used to simulate myofibroblast proliferation. FBA was employed to assess the metabolic fluxes of glycolysis during proliferation in the FB_N and FB_{TGF} models. Firstly, the bioavailability of glucose was increased by leaving

corresponding exchange reaction unconstrained (lower bound equals -1000). Then, FBA computed the maximal biomass rate. Flux rates for the glycolytic reactions in the FB_{TGF} model were then compared with those in the FB_N model. Next, FBA computed the maximal optimal flux for the collagen synthesis reaction in both models. FVA was used to assess alternate optimal solutions for a computed FBA with collagen synthesis as the objective function. FVA provides a range of fluxes (maximum and minimum flux) that a reaction can carry to achieve the same optimal value for the objective function. Then, flux span ratios were used to assess the essentiality of each reaction to the collagen production reaction. These flux spans and ratios were sorted in ascending order and those reactions with the smallest flux span or ratio were considered essential for the collagen synthesis. Single gene deletion was used to assess the essentiality of all genes in the model to collagen synthesis. This method provided a list of genes that impacted collagen synthesis and the calculated collagen synthesis rate for each gene that is deleted. To get the most essential genes out of that list, one can set a tolerance level (threshold) for the collagen synthesis rate. Next, genes whose deletion would generate a synthesis rate lower than threshold can be identified. The lower the threshold, the greater the essentiality of that gene to the objective reaction. Escher (King et al. 2015) was used to visualise the metabolic pathways and flux distributions from the model.

3.3 Results

3.3.1 Model generation

RNA-seq data were used to generate two separate models for the FB_{TGF} and $FB_{N.}$ These genome-scale models were constructed and refined using identical methods (as described in a previous section). The final version of the FB_N model accounts for 1673 genes, 2949 reactions and 2012 metabolites while the final FB_{TGF} model accounts for

1673 genes, 2909 reactions and 1994 metabolites. These models were made publicly available (GitHub Repository https://github.com/andreidob91/cobramodels). The reactions in the FB_N model are distributed over 81 metabolic processes (subsystems) while the reactions in the FB_N model account for 80 metabolic processes (Fig. 3.3 - 3.4). Interestingly, the only metabolic process that appears in the FB_N model and is not part of the FB_{TGF} model is histidine metabolism. In the FB_N model, histidine metabolism is represented by three reactions which involve histamine (catabolism of histamine). Differential data have been reported on the effects of histamine on fibroblast biology such as differentiation (Gailit et al. 2001, Lin et al. 2015) and although most reports suggest pro-fibrotic mechanisms (Kohyama et al. 2009), the effects of histamine may be context specific (Jordana et al. 1988). In support of a pro-fibrotic role for histamine, the presence of mast cells (which are the major source of histamine) has been noted in fibrotic lesions (Levick et al. 2018). However, in this case, the gene expression data and modelling work may suggest a negative effect of histamine on fibroblast differentiation.

The metabolites in both models are distributed over 9 compartments (cytosol, extracellular, Golgi apparatus, mitochondrial intermembrane space, lysosome, mitochondria, nucleus, endoplasmic reticulum and peroxisome). Figures 3.1 - 3.4 detail the reaction distribution and metabolite compartmentalisation in both the FB_N and FB_{TGF} model. Cytosol contains the greatest number of metabolites in the FB_{TGF} and FB_N models, followed by mitochondria. In terms of reaction distribution over metabolic processes, the highest number of reactions are extracellular transport reactions, followed by fatty acid oxidation.

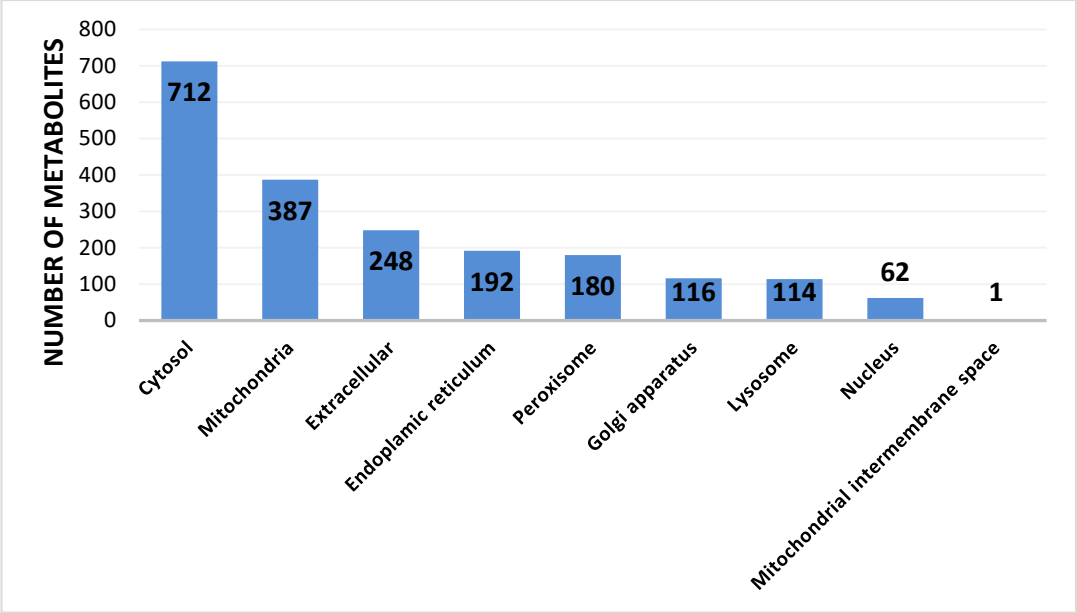
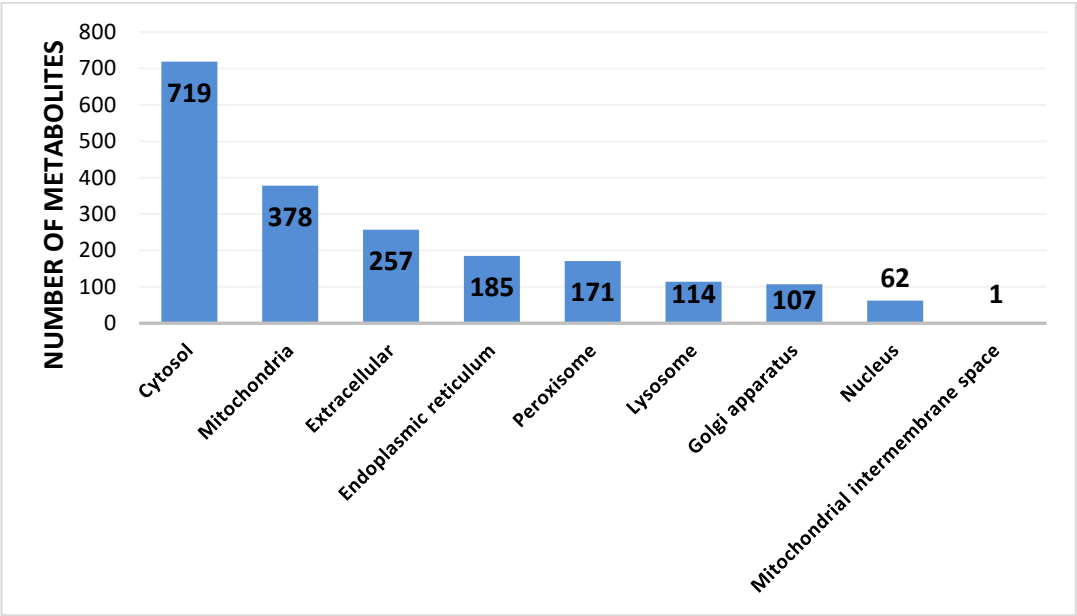


Figure 3.1 Compartmentalisation of metabolites in the FB_N model
This bar chart shows the cellular compartments in the FB_N model sorted by the number of metabolites in each compartment.



/

Figure 3.2 Compartmentalisation of metabolites in the FB_{TGF} model
This bar chart shows the cellular compartments in the FB_{TGF} model sorted by the number of metabolites in each compartment.

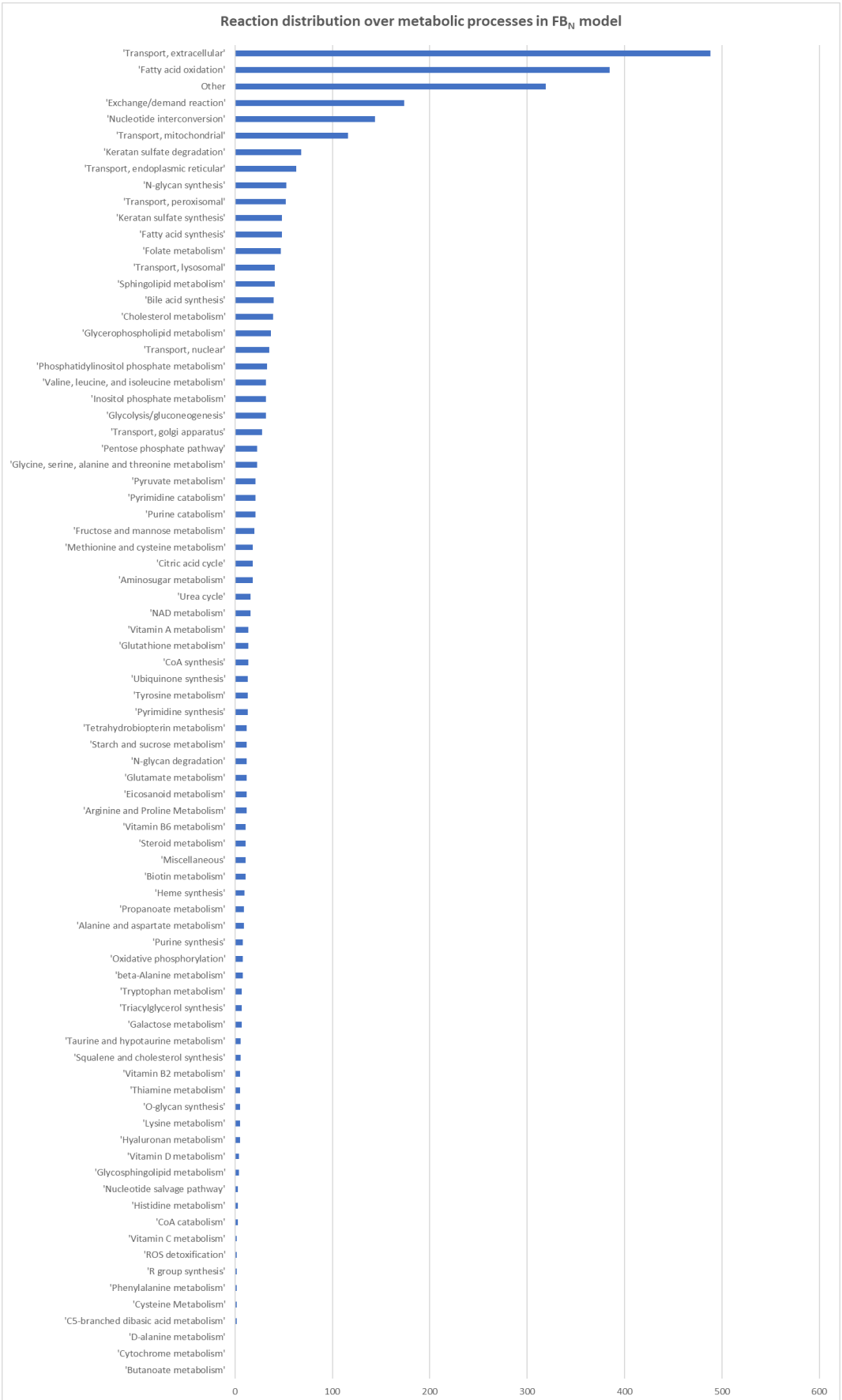


Figure 3.3 Reaction distribution over metabolic processes in the FB$_N$ model
This bar chart shows the metabolic processes in the FB$_N$ model sorted by the number of reactions in each metabolic process.

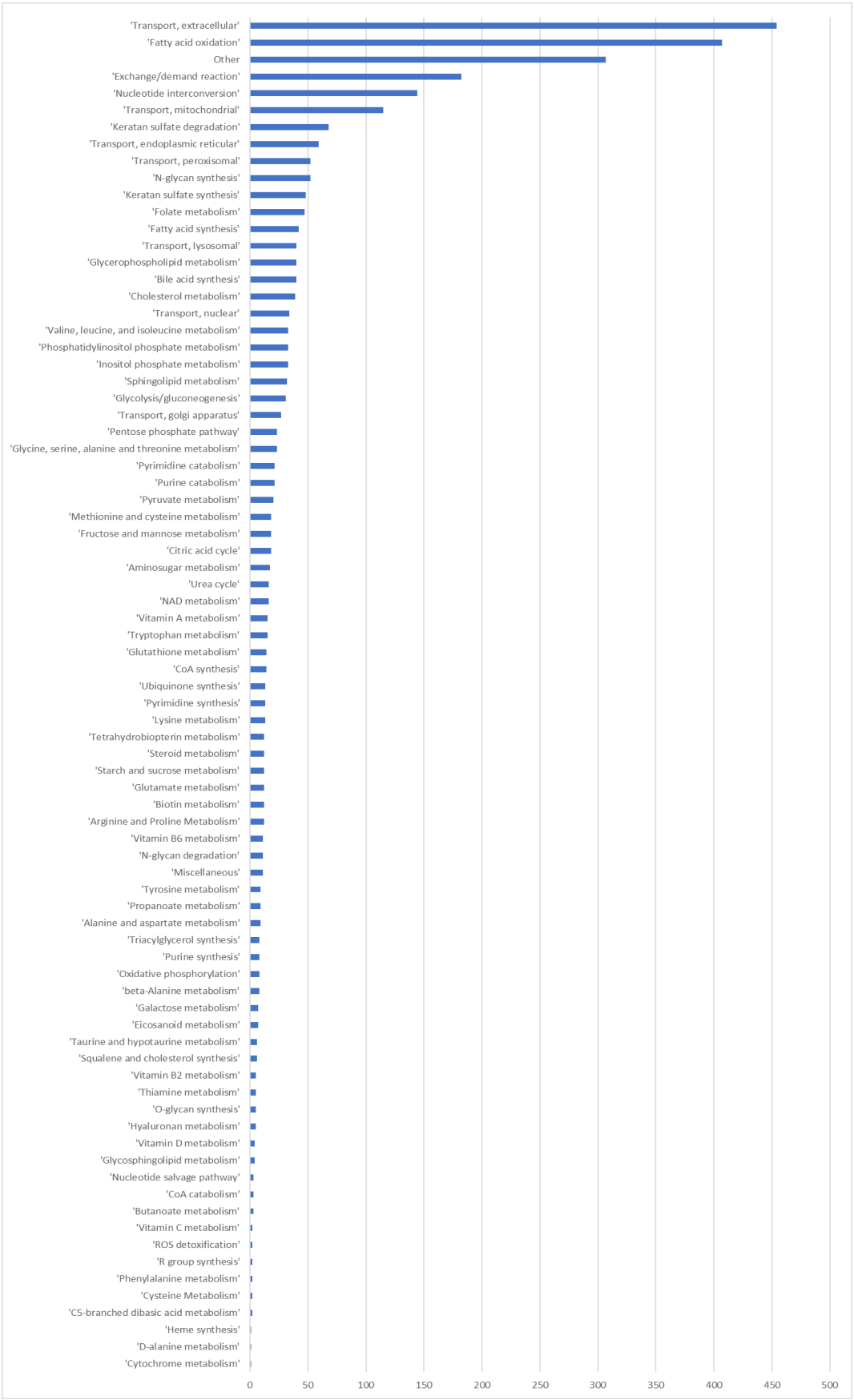


Figure 3.4 Reaction distribution over metabolic processes in the FB_{TGF} model
This bar chart shows the metabolic processes in the FB_{TGF} model sorted by the number of reactions in each metabolic process.

3.3.2 Flux balance analysis

FBA showed that the biomass flux rate was higher in the FB_{TGF} model, compared to the FB_N model (Table 3.1). Histidine was the first limiting AA in both models when they were optimised towards biomass production (Table 3.2-Table 3.3). Histidine is not synthesised internally by the model and is imported directly in the biomass reaction. Thus, the difference between the maximal biomass fluxes is mainly attributed to histidine uptake rate which is two times higher in the FB_{TGF} model compared to the FB_N model. If the lower bound for histidine uptake is doubled in the FB_N, the next limiting factor for biomass reaction is oxygen (followed by glucose, lysine etc.). When the same exchange constraints were imposed in the FB_{TGF} and FB_N models, the corresponding maximal fluxes for biomass production were equal and the models utilised very similar pathways to produce biomass.

The biomass objective function was further used to investigate metabolic reprogramming of FB_N and FB_{TGF} during proliferation, focusing on glycolysis. When optimising towards the biomass objective function (proliferation), there was a 2-fold increase in glucose uptake in the FB_{TGF} model compared to FB_N model. Furthermore, all rate-limiting glycolytic reactions such as HK, PFKM and PKM were predicted to have higher metabolic fluxes in the FB_{TGF} model cells compared to FB_N due to a higher glucose uptake rate. In consequence, the lactate production flux was higher in the FB_{TGF} model by 2-fold. *In vivo* studies have reported a 2.5-fold increase in lactate production for FB_{TGF} (Zhang et al. 2015). These results are in full agreement with the Warburg effect and the previously mentioned studies (Xie et al. 2015, Zank et al. 2018) which have reported that TGF-β stimulates glycolysis during myofibroblast proliferation in fibrotic diseases. Table 3.2 details these fluxes in both FB_{TGF} and FB_N models. It is important to point out that

these flux differences in the FB_{TGF} and FB_N reactions are determined by the differences in the exchange constraints.

Table 3.1 Maximal fluxes for biomass reaction in the FB_N and FB_{TGF} models

Biomass Production Rate (mmol gDW^{-1} h^{-1})	
FB_N	**FB_{TGF}**
0.013	0.027

**Flux rates were determined using Flux Balance Analysis*

Table 3.2 The lower bounds in the FB_N model vs. the actual uptake/secretion rates when optimising for biomass production

Exchange Reaction	Lb (max uptake rate mmol h^{-1} gDW^{-1})	Ub (max secretion rate mmol h^{-1} gDW^{-1})	Uptake rate (mmol gDW^{-1} h^{-1})	Secretion rate (mmol gDW^{-1} h^{-1})
Glc	-0.5280	0	-0.0863	N/A
Gln	-0.0290	1000	-0.0115	N/A
Arg	-0.0160	1000	-0.0086	N/A
Asn	-0.0041	1000	-0.0026	N/A
Asp	-0.0063	1000	-0.0054	N/A
His	-0.0032	1000	-0.0032	N/A
Leu	-0.0190	1000	-0.0087	N/A
Lys	-0.0120	1000	- 0.0074	N/A
Ile	-0.0130	1000	-0.0022	N/A
Met	-0.0049	1000	- 0.0020	N/A
Phe	-0.0058	1000	-0.022	N/A
Ser	-0.0370	1000	-0.0046	N/A
Thr	-0.0084	1000	-0.0026	N/A
Trp	-0.0015	1000	-7.32×10^{-4}	N/A
Tyr	-0.0046	1000	-0.0017	N/A
Val	-0.0140	1000	-0.0031	N/A
Lac	0	0.695	N/A	0.0038
O_2	-0.178	1000	-0.1227	N/A

**Flux rates were determined using Flux Balance Analysis*

Table 3.3 The lower bounds in the FB_{TGF} model vs. the actual uptake/secretion rates when optimising for biomass production

Exchange Reaction	Lb (max uptake rate mmol gDW^{-1} h^{-1})	Ub (max secretion rate mmol gDW^{-1} h^{-1})	Uptake rate (mmol gDW^{-1} h^{-1})	Secretion rate (mmol gDW^{-1} h^{-1})
Glc	-1.056	0	-0.2409	N/A
Gln	-0.058	1000	-0.0303	N/A
Arg	-0.032	1000	-0.0172	N/A
Asn	-0.0082	1000	-0.0052	N/A
Asp	-0.0126	1000	N/A	N/A
His	-0.0064	1000	-0.0064	N/A
Leu	-0.0380	1000	-0.0175	N/A
Lys	-0.0240	1000	-0.0148	N/A
Ile	-0.0260	1000	-0.0045	N/A
Met	-0.0098	1000	-0.0041	N/A
Phe	-0.0116	1000	-0.0043	N/A
Ser	-0.0740	1000	-0.0171	N/A
Thr	-0.0168	1000	-0.0053	N/A
Trp	-0.0030	1000	-0.0015	N/A
Tyr	-0.0920	1000	-0.0033	N/A
Val	-0.0280	1000	-0.0061	N/A
Lac	0	1.39	N/A	0.1390
O_2	-0.3560	1000	-0.3560	N/A

**Flux rates were determined using Flux Balance Analysis*

Table 3.4 Maximal flux rates for glycolytic reactions in the FB_N and FB_{TGF} models when optimising for biomass production

Reaction	Flux rate (mmol gDW^{-1} h^{-1})	
	FB_N	**FB_{TGF}**
Glucose Uptake	0.08	0.17
HK	0.08	0.17
GPI	0.03	0.07
PFKM	0.03	0.001
ALDOA	0.03	0.08
TPI	0.03	0.07
GAPDH	0.07	0.15
PGK	0.07	0.15
PGAM	0.07	0.15
ENO	0.07	0.15
PKM	0.07	0.15
Lactate Efflux	0.07	0.13

**Flux rates were determined using Flux Balance Analysis*

FBA results also showed a higher collagen production rate in the FB_{TGF} model compared to the FB_N model (Table 3.5). FB_{TGF} also display a higher collagen production *in vivo* compared to FB_N. The maximum collagen synthesis rate in the FB_{TGF} model was 0.000374 mmol gDW^{-1} h^{-1}, similarly to the *in vitro* rate of 0.000712 mmol gDW^{-1} h^{-1} (Illsley et al. 2000).

Oxygen, glutamine, aspartate, asparagine, isoleucine, threonine, valine and serine were hitting their uptake limits in the FB_N model when it was optimised for collagen production (Table 3.6). The lower bounds for these uptake rates were higher in the FB_{TGF}

model hence the increase in the collagen synthesis flux. Glycine was also required for collagen synthesis and no external uptake was considered for it. Thus, the maximisation of the collagen synthesis function is also dependent on the import/synthesis of serine which was converted to glycine through SHMT. When optimising for collagen synthesis, FBA also showed that the flux rates for PHGDH, PSPH, PSAT and SHMT1 were higher in the FB_{TGF} model compared to FB_N model (Table 3.8). Serine is imported in the model and then combines with the serine generated from the glycolysis-diverted PSPH reaction to produce glycine through SHMT reaction (Fig. 3.5 a-b). Owing to higher uptakes for serine and glucose, the fluxes corresponding to the *de novo* serine-glycine pathway are higher as well.

Table 3.5 Maximal flux rates for collagen synthesis rate in FB_N and FB_{TGF} models

Collagen Synthesis Rate (10^{-4} mmol gDW^{-1} h^{-1})	
FB_N	**FB_N**
1.86	3.74

**Flux rates were derived using Flux Balance Analysis*

With a higher glucose and oxygen availability, glycolysis and the TCA cycle also display higher fluxes in the FB_{TGF} model. Although in very small numbers, the fluxes corresponding to the Pentose Phosphate Pathway (PPP) are lower in the FB_{TGF} model compared to FB_N model. One significant difference between the models is related to the metabolic fate of pyruvate. With a higher glucose uptake, the cytosolic pyruvate accumulates and is converted to cytosolic lactate through LDHA. Forty six percent of lactate is secreted out of the model through the lactate exchange reaction. However, the amount of lactate exceeds the upper bound of the lactate exchange reaction (the exchange reaction is hitting its secretion limits) and the remaining 53% is transported to mitochondrial lactate and then converted to mitochondrial pyruvate through lactate dehydrogenase B (LDHB). Next, pyruvate dehydrogenase (PDH) converts mitochondrial

pyruvate to acetyl-CoA which replenishes the TCA cycle. In the FB_N model, the cytosolic pyruvate is converted to cytosolic lactate through LDHA and most of the lactate (97%) is secreted out of the model through the lactate exchange reaction. The remaining 3% is transported to mitochondria and converted to pyruvate. Mitochondrial pyruvate is then converted to oxaloacetate through pyruvate carboxylase (PC) reaction. Figure 3.5 (a-b) details these processes in greater detail.

Table 3.6 The lower bounds in the FB_N model vs. the actual uptake/secretion rates when optimising for collagen synthesis

Exchange reaction	Lb (max uptake rate mmol h^{-1} gDW^{-1})	Ub (max secretion rate mmol h^{-1} gDW^{-1})	Uptake Rate	Secretion Rate
Glc	-0.5280	0	-0.03811	N/A
Gln	-0.0290	1000	-0.0290	N/A
Arg	-0.0160	1000	-0.0133	N/A
Asn	-0.0041	1000	-0.0041	N/A
Asp	-0.0063	1000	-0.0063	N/A
His	-0.0032	1000	-0.0017	N/A
Leu	-0.0190	1000	-0.0090	N/A
Lys	-0.0120	1000	-0.0106	N/A
Ile	-0.0130	1000	-0.0130	N/A
Met	-0.0049	1000	-0.0024	N/A
Phe	-0.0058	1000	-0.050	N/A
Ser	-0.0370	1000	-0.0370	N/A
Thr	-0.0084	1000	-0.0084	N/A
Trp	-0.0015	1000	-0.0011	N/A
Tyr	-0.0046	1000	-0.0024	N/A
Val	-0.0140	1000	-0.0140	N/A
Lac	0	0.695	N/A	0.695
O_2	-0.178	1000	-0.35	N/A

**Flux rates were derived using Flux Balance Analysis*

Table 3.7 The lower bounds in the FB_{TGF} model vs. the actual uptake/secretion rates when optimising for collagen synthesis

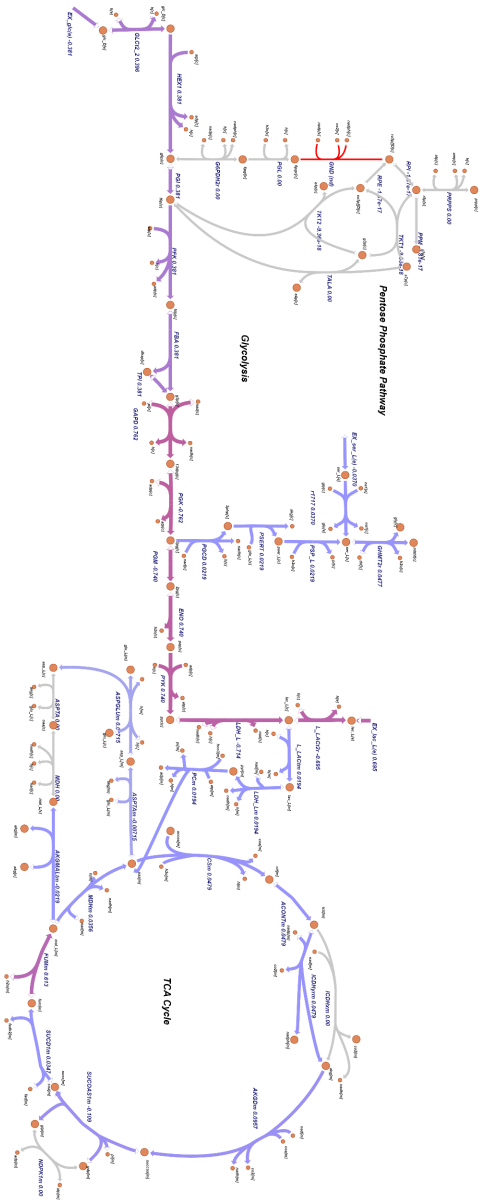
Exchange reaction	Lb (max uptake rate mmol h^{-1} gDW^{-1})	Ub (max secretion rate mmol h^{-1} gDW^{-1})	Uptake Rate	Secretion Rate
Glc	-1.056	0	0.44	N/A
Gln	-0.058	1000	-0.058	N/A
Arg	-0.032	1000	0.026	N/A
Asn	-0.0082	1000	-0.0082	N/A
Asp	-0.0126	1000	-0.0126	N/A
His	-0.0064	1000	-0.0034	N/A
Leu	-0.0380	1000	-0.0303	N/A
Lys	-0.0240	1000	-0.0213	N/A
Ile	-0.0260	1000	-0.0090	N/A
Met	-0.0098	1000	-0.0049	N/A
Phe	-0.0116	1000	-0.0101	N/A
Ser	-0.0740	1000	-0.0740	N/A
Thr	-0.0168	1000	-0.0168	N/A
Trp	-0.0030	1000	-0.0022	N/A
Tyr	-0.0920	1000	-0.0920	N/A
Val	-0.0280	1000	-0.0280	N/A
Lac	0	1.39	N/A	1.39
O_2	-0.3560	1000	-0.3560	N/A

**Flux rates were determined using Flux Balance Analysis*

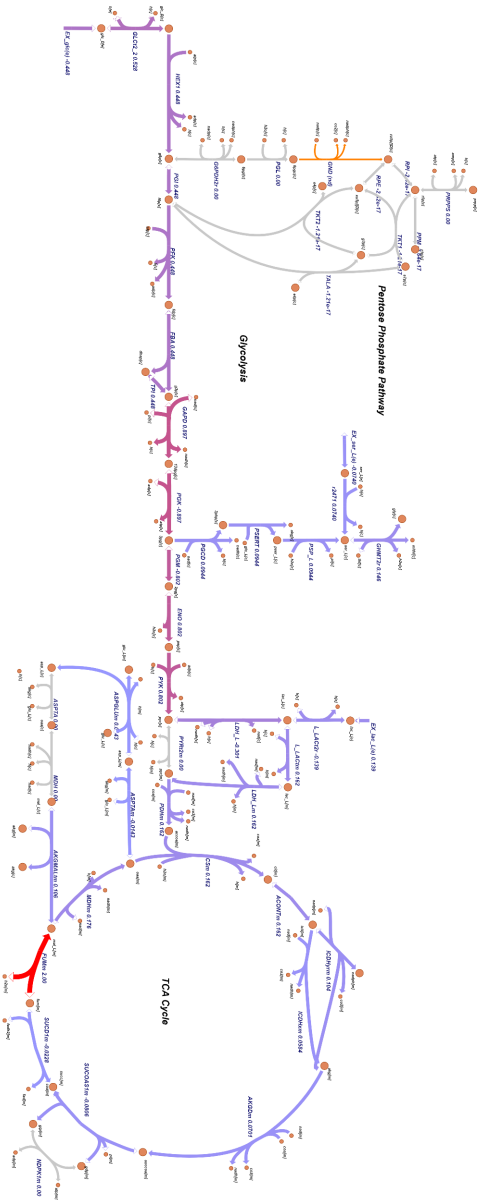
Table 3.8 Maximal flux rates for the reactions serine/glycine biosynthetic pathway when optimising for collagen production

Reaction	Flux rate (mmol gDW^{-1} h^{-1})	
	FB_N	**FB_{TGF}**
PHGDH	0.021	0.094
PSAT	0.021	0.094
PSPH	0.021	0.094
SHMT	0.047	0.146

**Flux rates were derived using Flux Balance Analysis*



(a)



(b)

Figure 3.5 The metabolic maps describing the glycolysis, pentose phosphate pathway and TCA cycle in the FBN (a) and FBTGF (b) models when collagen synthesis is optimised
Following its uptake, glucose is transported to cytosol where converted to pyruvate. Pyruvate replenishes the TCA cycle through different reactions: In the FBN model, pyruvate is converted to oxaloacetate through PC reaction to form oxaloacetate, while in the FBTGF model, pyruvate is oxidised via PDH reaction to form acetyl-CoA. The rest of pyruvate is converted to lactate and exported to the extracellular compartment through lactate exchange reaction. Due to higher uptake rates for glucose and oxygen, glycolysis and TCA cycle display higher fluxes in the FBTGF model. Although with very small fluxes, PPP is present in both model and displays lower fluxes in the FBTGF model. The de novo serine-glycine is present in both models and uses the 3-PG flux to produce serine which is further converted to glycine. Line thickness and colour (thin grey to bold red) are indicators of the flux values (low to high). The reactions which have no data in the model are coloured in orange. Escher (King et al. 2015) was used to generate the metabolic maps and visualise the flux distributions. Reaction and metabolite names were derived from Recon 2.2.

3.3.3 Flux variability analysis and flux spans

FVA and flux spans were used to rank reactions based on their essentiality to collagen synthesis. Four reactions from BCAA metabolism which generate precursors (NADH and FADH2) for oxidative phosphorylation appear in the top of the list. After this, electron transport chain reactions follow. Collagen production requires energy and oxidative phosphorylation is the primary source of ATP in the cell. Subsequently, the rest of the ATP needed for collagen synthesis is produced through glycolysis. Thus, TCA cycle and glycolytic reactions are also expected to be identified as essential. Table 3.9 details the results generated by flux spans.

Table 3.9 Reactions with smallest relative flux span when optimising FB_{TGF} model for collagen synthesis

Rank	Reaction	Flux span value	Metabolic Pathway
1	ACAD9	1.00	Valine, Leucine, and Isoleucine Metabolism
2	HIBADH	1.00	Valine, Leucine, and Isoleucine Metabolism
3	BCKDHA	1.00	Valine, Leucine, and Isoleucine Metabolism
4	ACAD8	1.62	Valine, Leucine, and Isoleucine Metabolism
5	COX	34.73	Oxidative Phosphorylation
6	NDUFA10	34.73	Oxidative Phosphorylation
7	ATP5F1B	34.73	Oxidative Phosphorylation
8	HK	735.37	Glycolysis
9	FH	999.07	Citric Acid Cycle
10	PKM	1246.21	Glycolysis

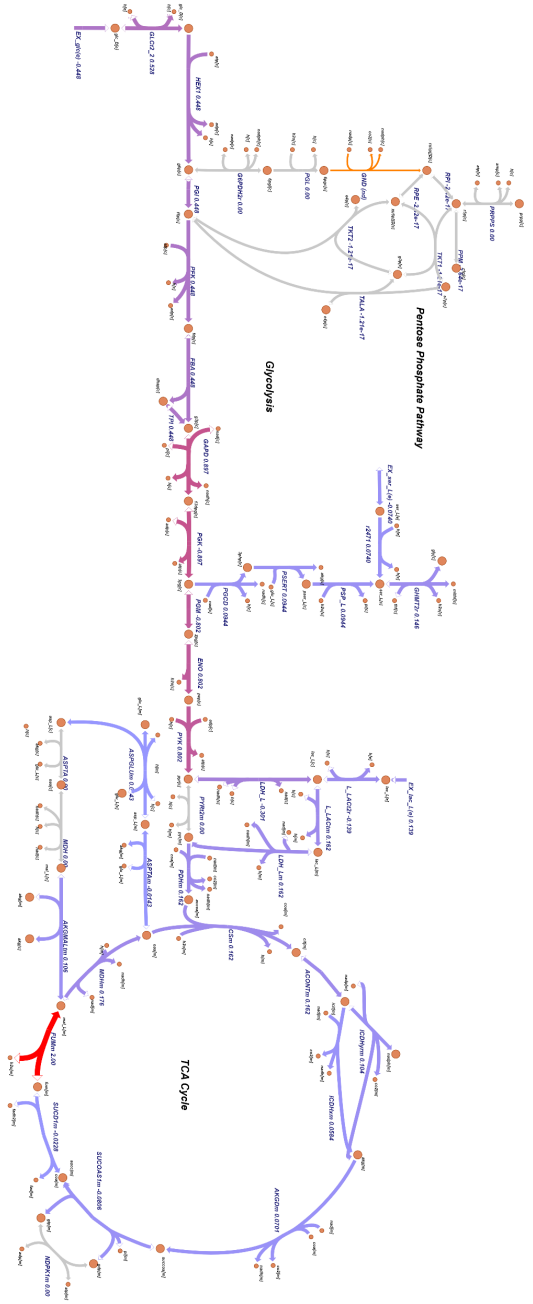
3.3.4 Gene essentiality analysis

Single *in silico* gene deletion identified a total of 581 genes that impacted (reduced) collagen synthesis rate. Most of them only had a minor effect on the collagen production reaction when deleted (Fig. 3.7). An initial threshold of zero was set for the synthesis

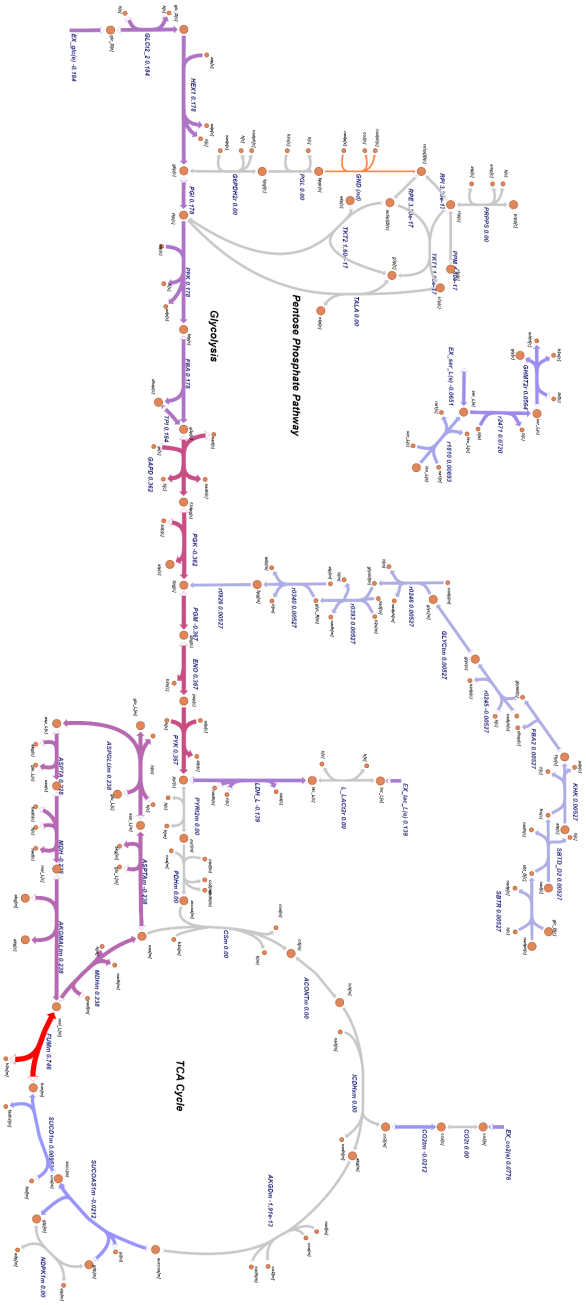
rate of collagen. Deletion of PYCR1 and SHTM1 results in a collagen synthesis rate of zero. Based on the current analysis, these genes are essential for collagen synthesis. The first one is an encoding of the enzyme that catalyses the last step of proline biosynthesis. The proline uptake rate was not set in the model and a good proportion of a collagen molecule is comprised of proline (23%). Furthermore, proline is used in used in many AA transport reactions hence this gene becomes essential for collagen synthesis. This fact is in agreement with the result generated by FVA which was discussed in the previous section. The second gene, SHTM1, encodes the cytosolic enzyme that catalyses the final step of glycine synthesis from serine. With no import of glycine in the model, the glycine synthesis is dependent on the SHMT gene which converts serine to glycine. Next, the threshold was progressively increased until it reached about half of the initial collagen synthesis rate. In this case, four more genes were identified: CS, PHGDH, PSPH, PSP (Table 3.10). This implies that deletion of these genes reduces the maximal collagen production rate by about 50%. CS encodes the rate-limiting enzyme catalysing the synthesis of citrate from oxaloacetate and acetyl-CoA in the first step of TCA cycle. The other three genes, PHGDH, PSPH and PSAT encode the enzymes catalysing the first three steps in the *de novo* serine-glycine pathway.

When CS was deleted and the model optimised for collagen production, the TCA cycle was reduced, but not entirely supressed. CS catalyses the initial reaction of the TCA cycle in which citrate is produced from acetyl-CoA and oxaloacetate. With the citrate synthase (CS) reaction being switched off, the next two reactions in the cycle, aconitate hydratase (ACO2) and isocitrate dehydrogenase (IDH2) were also supressed. 2-oxoglutarate dehydrogenase (OGDH) carried on with a very small flux as α-ketoglutarate was produced through alternative pathways in very small quantities. In this scenario, the

entire amount of succinyl-CoA was produced through an alternative but less efficient pathway involving the metabolism of valine. Hence the amount of ATP produced anaerobically was reduced. Under normal circumstances, 10.5% of the PGK flux is diverted into the *de novo* serine-glycine pathway via 3-PG and used to synthesise serine which is further converted to glycine. When CS is switched off, the entire flux for PGK is used to produce ATP and nothing is diverted into the *de novo* serine-glycine pathway. Furthermore, glycolysis continues as 3-PG is produced through an alternative pathway which includes the use of fructose and mannose metabolism. Thus, with no serine produced internally, the production of glycine relies entirely on the amount of serine which is imported through the exchange reaction hence the reduction in collagen synthesis (Fig. 3.6 a-b).



(a)



(b)

Figure 3.6 The metabolic maps describing the glycolysis, de novo serine-glycine pathway and TCA cycle in the FBTGF model with CS reaction being on (a) and off (b) when optimising for collagen synthesis

Following the deletion of CS gene, ACO2 and IDH3 reactions are stopped. In this framework, the flux corresponding to reaction is used to produce ATP and the serine/glycine synthesis relies only on the import of serine from the extracellular compartment. In consequence, the synthesis of collagen is reduced. At the same time, glycolysis carries on as 3-PG is produced through an alternative pathway. Line thickness and colour (thin grey to bold red) are indicators of the flux values (low to high). The reactions which have no data in the model are coloured in orange. Escher (King et al. 2015) was used to generate the metabolic maps and visualise the flux distributions. Reaction and metabolite names were derived from Recon 2.2.

A question that could arise is why mitochondrial CS was the only TCA cycle enzyme that became critical to collagen synthesis? One answer could be that Recon 2.2 does not provide a complementary gene for the mitochondrial CS or alternative routes to produce the necessary amount of α-ketoglutarate to maintain the initial flux of the collagen synthesis. When the other TCA cycle genes were subjected to single gene deletion, the model was able to find alternative genes/enzymes to maintain the TCA cycle running and produce enough α-ketoglutarate which guaranteed the optimal production of glycine through the *de novo* serine-glycine pathway. For instance, when the cytosolic aconitase (ACO1) was supressed, mitochondrial aconitase (ACO2) was used by the model to maintain the initial flux and vice-versa. It is important to point out that a combinatorial deletion of these genes (ACO1 and ACO2) would be critical to collagen synthesis reducing it to zero flux rate. Similarly, when the $NADP^+$-dependent isocitrate dehydrogenase (IDH2) was supressed, the NAD^+-dependent isocitrate dehydrogenase (IDH3) was used to produce α-ketoglutarate and the initial flux was maintained. When OGDH was deleted, dihydrolipoyllysine-residue succinyltransferase (DLST) was used to produce the necessary amount of succinyl-CoA and maintain the initial flux. When succinyl-CoA ligase (SUCLG1) was deleted, succinic semialdehyde dehydrogenase ALD5A1 was used to produce succinate. The deletion of succinate dehydrogenase (SDHA) and fumarate hydratase (FH) enabled the use of internal metabolic routes to produce fumarate and malate and maintain the initial collagen synthesis flux. Finally, when the mitochondrial malate dehydrogenase (MDH2) was deleted, the cytoplasmic malate dehydrogenase (MDH1) was used to maintain the initial flux.

Single gene deletion in the FB_{TGF} model also identified PHGDH, PSPH, PSAT as critical genes for collagen production. Furthermore, SHMT was identified to be essential for

collagen production. These genes are encoding those enzymes that catalyse reactions involved in glycine and serine *de novo* synthesis. The analysis for the FB_{TGF} model provided evidence for *de novo* serine-glycine biosynthetic pathway, identifying the correct reactions, enzymes and genes involved in this process and their essentiality for collagen production. This is an important result as TGF-β appears to play a vital role in regulating the gene expression of the enzymes involved in the serine *de novo* synthesis. The importance of this pathway to collagen synthesis will be further discussed in section 3.4.

Table 3.10 Single gene deletion results in the FB_{TGF} model

Deleted Gene	***Collagen Synthesis Rate (mmol gDW^{-1} h^{-1} x 10^{-4})***	***% reduction in the max collagen synthesis rate***
SHMT1	*0*	*100%*
PYCR1	*0*	*100%*
CS	*1.44*	*61%*
PSPH	*1.64*	*56%*
PHGDH	*1.64*	*56%*
PSAT1	*1.64*	*56%*

**Deletion of SHMT1 and PYCR1 resulted in a collagen synthesis rate of zero. Deletion of CS reduce the collagen synthesis rate by 61%, while deletion of PSPH, PHGDH and PSAT1 reduced the collagen synthesis rate by 56%.*

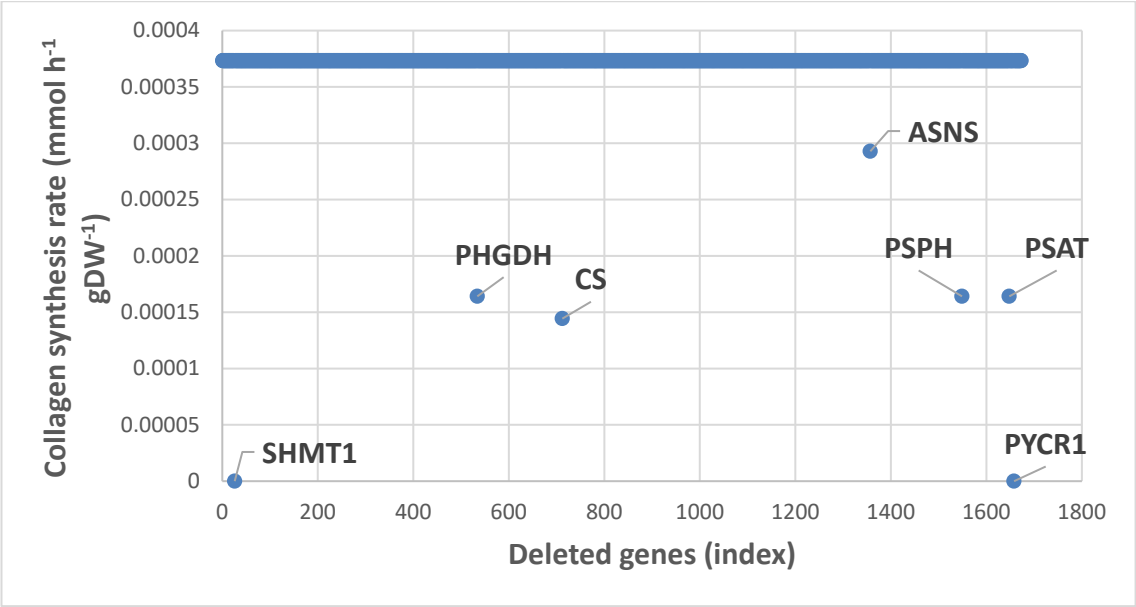


Figure 3.7 Gene essentiality analysis in the FB_{TGF} model
The collagen synthesis rate after single gene deletion (Y axis) is plotted against the deleted genes (X axis). Deletion of most genes had a minor effect on the collagen synthesis reactions. These genes are plotted very near to the collagen synthesis rate (3.74×10^{-4} mmol gDW^{-1} h^{-1}). Deletion of SHMT1 and PYCR1 genes generated a zero collagen synthesis rate and these genes were considered essential for the objective function. Deletion of PHGDH, PSAT1 and PSPH reduced the collagen synthesis rate to 1.64×10^{-4} mmol gDW^{-1} h^{-1} while deletion of CS gene reduce it to 1.44×10^{-4} mmol gDW^{-1} h^{-1}. These genes were considered critical to collagen synthesis.

In order to make sure that the results described earlier were not influenced by the fact that for certain AAs the uptake rates were not provided, an analysis was performed with these AAs being exchanged (influx was allowed). As was mentioned in the methods section, the alanine, proline, glycine, cysteine AAs were reported not to be consumed by the cells and have their corresponding lower bound equal to zero. For cysteine and glutamate, the exchange reactions were not picked up in the initial model building process by FASTCORMICS. As outlined in section 3.2.1, if these exchanges were not a part of the core set, they were likely to be removed and the model was able to produce them internally. To perform this analysis, the corresponding exchange reactions for glutamate and cysteine were manually added to the model.

Given that there were no uptake rates for any of these AAs in the literature data source, the lower bound for each of these AAs was calculated as the mean of the lower bounds

for all consumed AAs. In this framework, the model was optimised for collagen production. No significant changes were observed. However, there was a slight decrease in the percentage by which collagen synthesis dropped when PHGDH, PSPH, PSAT were deleted (from 56.1% to 39.5%). With glycine and proline uptake in place, the essentiality of the serine-glycine metabolic pathway was reconfirmed, and the same genes were predicted as critical/essential by the GEA with the exception of PYCR1, which was responsible to produce proline in the absence of an uptake rate for this AA. It is important to point out that the uptake of glycine was not able to fully compensate for the amount of glycine needed by collagen synthesis and the serine-glycine pathway remains essential to the production of collagen.

3.3.5 Differential expression analysis

Differential expression analysis was used to investigate the change in the expression levels under "control" and "TGF-β-stimulated" conditions. In this way, the results generated from FBA, FBA-based GEA and FVA could be verified using a method which is entirely based on "omics" data. Initially, the analysis produced a table containing 22110 genes with non-zero total read count. Out of these, 4930 were up-regulated (p-adjusted value < 0.05) and 4813 were down-regulated (p-adjusted value < 0.05). The analysis also identified 1 outlier and 4139 genes with a low count number.

Table 3.11 DEA results

Total Genes	LFC>0 (up) P<0.05	LFC<0 P<0.05	Outliers	Low count
22120	4930	4813	1	4139
	22% total	22% total	0.045% total	19% total

**DEA results showing the total number of: genes analysed, up-regulated genes, down-regulated genes, outliers and genes with low count.*

Next, a volcano plot was used to display the gene fold-change versus the statistical significance. Thus, genes involved in the glycine/serine pathway and the rate-limiting glycolytic genes were investigated using the volcano plot. The results showed that genes

coding for the enzymes involved in the glycine/serine pathway are up-regulated under TGF-β stimulation with the exception of SHMT1 whose fold-change was not statistically significant. However, SHMT2 was picked up as significantly up-regulated under TGF-β stimulation with a log2 fold-change of 2.01 and a p-value of zero. Thus, these results are consistent with FBA and GEA which showed that these genes are essential/critical for collagen production and the corresponding reactions have higher fluxes in the FB_{TGF} model. Next, the genes coding for the rate-limiting glycolytic enzymes were also displayed in the volcano. The results showed no significant difference in gene expression for HK1, PKM, PFKM under TGF- β stimulation. HK1 were placed in the non-significant area of the plot while PFKM had a small p-value but a relatively small fold-change. These results are not consistent with the FBA results which showed higher fluxes for HK1, PKM and PFKM in the FB_{TGF} model and with the *in vivo* studies which reported an upregulation of these enzymes under TGF- β stimulation (Fig. 3.8).

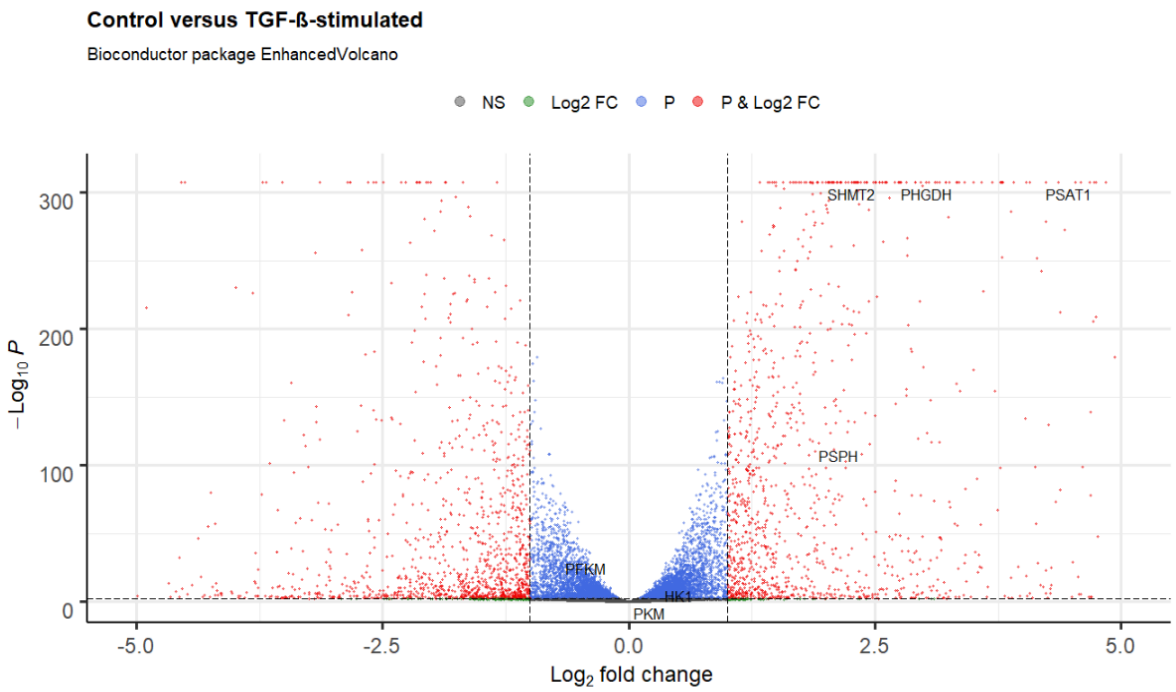


Figure 3.8 Volcano plot displaying significance vs fold-change under TGF- β stimulation
The X-axis represents the log2 fold-change while the Y-axis shows the negative log10 of the p-value. The two vertical lines correspond to the up-regulated and down-regulated fold-change thresholds (-5,5) while the horizontal line represents the p-value threshold (0.05). Each dot in the figure represents a single gene. Red colour represents either an up-regulation (right) or a down-regulation of the gene expression under TGF- β stimulation (high fold change and low p-value). Blue corresponds to low p-value but small fold-change, while green corresponds to high p-value and low fold-change. Grey represents the non-significant (NS) genes with high p-value ad low fold-change.

3.4 Discussion

Myofibroblast cells play a crucial role in supporting the wound healing process. They contract the wound and facilitate re-epithelisation. Under normal circumstances, myofibroblasts disappear through apoptosis. However, pathological cases such as fibrotic diseases are characterised by myofibroblast persistence, altered ECM and excessive collagen deposition. IPF is an idiopathic fibrotic disease that causes scarring in the lungs over time. The prognosis for IPF patients is poor and the available treatment can only relieve some symptoms. Some studies have reported that myofibroblasts metabolic reprogramming occurs in IPF (Xie et al. 2015, Zank et al. 2018). Although metabolic reprogramming has been thoroughly studied in cancer cells and glycolysis is

considered a drug target, this phenomenon has not been fully understood in fibrosis. TGF-β-induced myofibroblast differentiation has been associated with augmented glycolysis within the cell. Furthermore, in IPF patients, the *de novo* serine-glycine biosynthetic pathway and the enzymes catalysing the reactions in this pathway are reported to be up-regulated in IPF. Collagen has a unique AA build-up with about one third of collagen being glycine. Serine is diverted from glycolysis via 3-PG and then converted to glycine in a three-step enzymatic reaction. Thus, the *de novo* serine-glycine pathway constitutes a major metabolic pathway that is required for collagen synthesis. In IPF, TGF-β was reported to not only stimulate glycolysis, but also to upregulate the enzymes catalysing the reactions in the serine-glycine pathway (Nigdelioglu et al. 2016).

Using the global human metabolic reconstruction, FASTCORMICS algorithm and transcriptomics data, two genome-scale metabolic models for the FB_N and FB_{TGF} cells were constructed. Initially, transcriptomics data generated from microarrays were used to generate a model for the myofibroblast cell. This was laser capture microdissection of human idiopathic pulmonary fibrosis lung samples collecting fibroblastic foci. The advantage of using microarray data was that of having a myofibroblast-specific gene signature in IPF. However, the limitation was that only a small amount of RNA was obtained, which resulted in an incomplete representation of gene expression in the cells. Given this limitation, transcriptomics data generated by RNA-seq from cultured fibroblasts was used as cultured fibroblasts provided sufficient RNA to perform deep RNA-seq. Human primary fibroblast cells from healthy volunteers were stimulated with TGF-β to induce myofibroblast phenotype. FB_{TGF} is thought to more closely represent myofibroblast phenotype and IPF. These two data sets (FB_N and FB_{TGF}) gave us the

opportunity to produce two models and compare them using constraint-based methods.

The model aimed to investigate whether metabolic reprogramming occurs in myofibroblasts using constraint-based methods such as FBA, FVA and GEA. Initially it was checked whether FB_{TGF} model exhibits augmented glycolysis. The results showed that glucose uptake was higher in the FB_{TGF} model when compared to the FB_N model. Next, the rate-limiting reactions in glycolysis were found to display a higher maximal flux rate in the FB_{TGF} model when compared to FB_N model. Lactate secretion was also higher in the FB_{TGF} model when compared to the FB_N model. However, these differences in the fluxes were determined by the differences in the uptake rates for histidine and valine. The results from DEA did not confirm the hypothesis that the genes coding for rate-limiting enzymes in glycolysis were up-regulated under TGF-β stimulation. However, this was confirmed by the aforementioned *in vivo* studies which may suggest that biological hypotheses cannot be gleaned from "omics" data alone.

Next, the collagen production reaction was chosen as an objective function for the models and FBA was computed to calculate maximum collagen production rate. The model reported a higher collagen production rate in the FB_{TGF} model compared to FB_N model and displayed higher fluxes for the glycolysis and TCA cycle in the FB_{TGF} model. In terms of the structure, the models use very similar pathways to produce collagen and the difference between the fluxes in the glycolysis and TCA cycle is attributed to the difference between uptake/secretion rates. However, one noticeable difference is the metabolic fate of pyruvate. In the FB_N model, pyruvate is converted to oxaloacetate through PC, while in the FB_{TGF} model pyruvate is converted to acetyl-CoA through PDH.

Having compared both models, the FB_{TGF} model was analysed in more detail to identify potential drug targets. Following this, other FBA-based tools were employed to further assess the essentiality of certain reactions and genes to collagen production. FVA and GEA were used to identify which enzymes and genes are essential to collagen production. GEA identified PHGDH, PSPH and PSAT as critical genes to collagen production. SHMT was even found essential to collagen synthesis, its deletion generating zero collagen production rate. This result was encouraging as the model correctly predicted the fact that *de novo* serine-glycine pathway is important, even essential, to collagen production. It may be argued that this result could be influenced by the fact that glycine influx was not permitted in the model. To verify this hypothesis, the glycine import was temporarily allowed, and the same simulations were performed. GEA still identified SHMT as an essential gene while PHGDH, PSPH and PSPH were found critical to collagen production.

Lastly, the fluxes of the key reactions involved in this pathway were evaluated. FBA was used to compare the fluxes in the FB_N and FB_{TGF} models. The results showed that flux rates for the reactions involved in serine/glycine biosynthesis were higher in the FB_{TGF} model. The results were in agreement with an *in vivo* study published in 2016 (Nigdelioglu et al. 2016). This paper describes several key roles of TGF-β in stimulating glycolysis, synthesis of glycine, serine and collagen in IPF patients. Firstly, the study reported that TGF-β upregulated glycolysis. However, transcriptional regulation of TGF-β was not influenced by glycolysis. Thus, it was suggested that a glycolysis-diverted pathway may impact the TGF-β-stimulated collagen synthesis. Then, this paper aimed to investigate whether biosynthesis of serine/glycine was augmented in the presence of TGF-β. When the FB_N and FB_{TGF} cells were compared, it was reported that the gene

expression of the enzymes involved in serine/glycine biosynthesis (PHGDH, PSPH, PSAT, SHMT) was upregulated in the FB_{TGF}. This is in full agreement with the results from FBA when the fluxes corresponding to these enzymes in both the FB_{TGF} and FB_N models were compared. Furthermore, this review also suggested that *de novo* synthesis of serine and its conversion to glycine were required for TGF-β-stimulated collagen production. *In vivo* gene knockout of PHGDH and SHMT2 led to lower collagen production when treated with serine and glycine in culture, compared to control. Moreover, TGF-β-stimulated collagen production was completely disabled when cells (with knockdowns of PHGDH and SHMT2) were not cultured with glycine or serine. In other words, PHGDH and SHMT2 were identified as essential genes. Thus, *in silico* single gene deletion generated similar results to *in vivo* gene knockout, with exception of PHGDH, which was not identified as essential gene to collagen synthesis. GEA only identified PHGDH as critical gene, whose deletion reduced the collagen synthesis by about 50%. In a separate analysis, DEA showed that under TGF-β stimulation, the gene expression levels for PHGDH, PSAT, PSPH, and SHMT were significantly up-regulated. This result is in agreement with the results generated from FBA which showed higher fluxes for the corresponding enzymes and can act as a confirmation for the essentiality of this metabolic pathway to collagen production in FB_{TGF}.

TGF-β-stimulated myofibroblast differentiation generates a cascade of perturbations of metabolic reactions, gene regulatory circuits and signalling pathways. Although genome-scale models have evolved considerably in recent years, it is impossible at the present time for such a model to capture all of these genetic and metabolic perturbations. This is mainly because of the lack of data regarding kinetic, metabolic parameters and enzymatic processes. For instance, even if all the enzymatic parameters

were determined *in vitro*, little is known about their post-translational modifications (Pfau et al. 2011). These models capture the GPR (gene-protein-reaction) relationship, rely almost entirely on stoichiometry and can predict steady-state flux distributions of the metabolic reactions. Further limitations of genome-scale metabolic models are discussed in section 5.2. However, these models can successfully identify critical metabolic pathways and analyse the effect of certain genetic perturbations on these pathways. Even though they do not capture the dynamics of metabolism in cells, they represent a very useful map of metabolism. As is mentioned earlier in the thesis, metabolic reprogramming in fibrosis has been less studied *in vivo*. The model confirmed the results from *in vivo* studies which dealt with augmented glycolysis and excessive collagen synthesis in IPF. Furthermore, the model confirmed the importance of the serine-glycine biosynthetic pathway for collagen production and the up-regulation of the enzymes involved in this pathway. These four enzymes can therefore be considered potential drug targets in IPF. To the best of the author's knowledge, this pathway has only recently been considered as potential drug target and only a few studies have investigated it in detail (Nigdelioglu et al. 2016, Selvarajah et al. 2019).

In summary, the results published here suggested that TGF-β stimulates cell proliferation and collagen synthesis. The FBA results also suggested that FB_{TGF} display augmented glycolysis during proliferation which is not confirmed by DEA but in full agreement with the already mentioned *in vivo* studies. Lastly, the simulations showed that *de novo* synthesis of serine and its conversion to glycine constitutes a critical metabolic pathway for collagen synthesis in FB_{TGF} which is confirmed by both DEA and *in vivo* experiments. Overall, it can be concluded that the use of genome-scale metabolic models combined with constraint-based analysis methods is likely to be an important

tool in contributing to the design of therapeutic targets against many conditions including fibrosis and IPF in particular.

Chapter 4: Reconstruction and analysis of skeletal muscle cell using constraint-based methods

Having established the CBMs for FB_{TGF} and FB_N, the next step was to extend the modelling utility to another contractile cell. This chapter describes a metabolic model for skeletal muscle which was constructed and analysed using constraint-based methods. Model analysis focused on investigating the effects of hypoxia and AA supplementation on protein synthesis rate in the skeletal muscle cell.

4.1 Introduction

In response to physical activity (specifically resistance exercise) and nutrient availability, muscle hypertrophy occurs. Myofibres increase in size, driven by a net protein balance. However, as was outlined in the introductory chapter, in humans, muscle hypertrophy is determined by an increase in protein synthesis rather than a decrease in protein breakdown. Muscle atrophy, which occurs in conditions such as immobilisation or aging, is associated with a reduction in protein synthesis (de Boer et al. 2007). The loss of muscle mass can also occur in mountain climbers, specifically for lowlanders sojourning at high altitudes (Boyer et al. 1984). Although hypoxia has been reported to blunt the increase in myofibrillar protein synthesis rate after resistance exercise (Etheridge et al. 2011), the mechanisms leading to the loss of muscle mass at high altitude remain unclear and the impact of high-altitude hypoxia on the protein synthesis rate has not yet been fully understood (Holm et al. 2010). In this chapter, the skeletal muscle cell's metabolic response to hypoxia was investigated using CBM methods with a focus on protein synthesis rate. By contrast to hypoxia, dietary supplementation with certain AAs stimulates protein synthesis rate and may aid the growth of muscle induced by exercise (O'Bryan et al. 2019), yet aspects of this process remain unclear. Hence another aim was

to investigate whether supplementation with certain AAs *in silico* would have the same effect on protein synthesis as during *in vivo* experiments and to identify novel combinations of AAs that would stimulate protein synthesis.

4.2 Methods

In this section, the components that are appropriate for both hypoxia and AA supplementation are provided, and where appropriate, it is made clear to which specific process a component of the methods pertains.

4.2.1 Initial model generation

Publicly available RNA-seq transcriptomics data for skeletal muscle tissue were downloaded from (ArrayExpress https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836). RNA-seq analysis was performed as in (Uhlén et al. 2015). Ensembl IDs were converted to corresponding Entrez Gene IDs using BiomaRt package in R. A cut-off value of 1 was chosen based on sensitivity analysis for the TPM values. Those genes with a TPM greater than 1 were called present and assigned 1, while those with TPM less than 1 were called absent and assigned -1. Apart from that, the model was built using almost identical methods as for FB_{TGF} and FB_{N} models. The FASTCORMICS workflow and GPRs in Recon 2.2 were used to generate a context-specific model for the skeletal muscle cell. COBRA Toolbox and MATLAB were used in this step.

4.2.2 Manual refinement

A manual refinement of the model was performed as described in section 2.2.6. The reactions for actin, myosin and acto-myosin complex building were added to the essential reactions set and “forced” into the model. The exchange reactions for proline, threonine, tryptophan, histidine and phenylalanine, were manually added to the model

as they were not picked up by FASTCORMICS in the initial model. Out of these reactions, proline is the only non-essential AA. However, it was reported to be imported in the muscle cell, thus the import of proline was made available and an uptake rate was considered. As it was mentioned earlier in the paper, the goal was to create a data-driven model which is consistent with the literature, rather than a model based entirely on omics data. The uptake/secretion constraints for AAs, oxygen glucose, lactate and lipids were derived from literature: oxygen, glucose, lactate, palmitate (Turcotte et al. 1998), AAs (Baños et al. 1973). Rat muscle data was used in this case. The chosen unit for flux data was μmol min^{-1} g^{-1} muscle. This is a common unit for measurements performed *in vitro* on muscle tissue. Where necessary, transformations were performed to obtain this unit.

4.2.3 Modelling assumptions

FBA and FVA were the principal constraint-based modelling tools that were used for all the simulations. Before we performed these simulations, a couple of assumptions were made: Firstly, the model was considered to be in steady state. The reason for this assumption is related to basic constraint-based modelling approaches and was further discussed in the Chapter 1. Secondly, the contractile unit was considered to be made up of the two major proteins actin and myosin to form Acto-Myosin complex. Troponin and tropomyosin were not included in the reaction. And finally, it was assumed that the skeletal muscle cell has already all the hormones available to facilitate muscle growth.

4.2.4 Model analysis

4.2.4.1 Studying the Effects of Hypoxia on Protein Synthesis

Using FBA, the maximal flux rate of myofibrillar protein synthesis was evaluated while progressively decreasing the oxygen uptake rate. To represent hypoxia in CBM, the

oxygen uptake rate is usually reduced to 10% of the normal uptake rate (Eyassu et al. 2017). Hence oxygen consumption was reduced to 10% of its normal value and was progressively dropped to 0% (anoxia) to investigate the impact of oxygen reduction on the objective function (protein synthesis rate). As certain enzymes were reported to be either up-regulated or inhibited in hypoxia, FBA was also used to evaluate the fluxes of reactions catalysed by these enzymes under normoxic and hypoxic conditions and investigate the metabolic pathways leading to these changes. Finally, flux spans and FVA were used to assess the role of certain reactions to hypoxic performance of the cell. As glucose uptake is supressed when fatty acids are consumed, these investigations were initially performed without fatty acid uptake so that glycolytic fluxes under normoxia and hypoxia can be compared.

4.2.4.2 Studying the effects of AA supplementation on protein synthesis

Using FBA, the effect of single EAA supplementation on the maximal protein synthesis rate was assessed. The lower bound of each EAA was relaxed (one at a time) and the Acto-Myosin building reaction was optimised. In this framework, “relaxed” means that the lower bound for the uptake rate is set to -1000. Next, using the same methods, the effect of BCAA (leucine, isoleucine, and valine) supplementation was evaluated. This time, the constraints were “relaxed” simultaneously for each of the three BCAAs to ensure they are supplemented in combination. Further, the effects of multiple supplementation (with all possible combinations of EAAs) on protein synthesis rate were evaluated. And finally, the results were compared with literature data. It is important to mention that prior to all simulations the lower bounds were considered equal to -1000 and for the non-essential AAs (influx was allowed). For multiple EAA supplementation a “Power Set” containing all EAA exchange reactions was generated. In mathematics, a

power set contains all the subsets of a given set A including the empty set and the initial set itself. For instance, for a set A = {a,b,c} the Power Set, denoted as P(A), is {{Ø};{a};{b};{c};{a,b};{a,c};{b,c};{a,b,c}}. If a given set contains n elements the number of elements in the power set is 2^n. For 9 EAAs the Power Set will contain 512 combinations including the empty set. An algorithm was developed to relax the constraints in all possible combinations for the EAAs exchange reactions and sort all combinations in ascending order based on their impact on the protein synthesis rate. MATLAB was used in this step.

4.3 Results

This section divides the results into three subsections: a general model description, the evaluation of the effects of hypoxia on skeletal cell metabolism and protein synthesis rate and the evaluation of the impact of AA supplementation on protein synthesis rate in skeletal muscle.

4.3.1 Model description

The final version of the skeletal muscle model was made publicly available (GitHub Repository https://github.com/andreidob91/cobramodels). It accounts for 1673 genes and 1368 reactions. These reactions account for 65 metabolic functions (subsystems). The model also contains 910 metabolites distributed over 9 compartments. Figures 4.1 and 4.2 detail the reaction and metabolite compartmentalisation in the model. Cytosol contains the greatest number of metabolites in the model, followed by mitochondria. In terms of reaction distribution over metabolic processes, the highest number of reactions are extracellular transport reactions, followed by fatty acid oxidation.

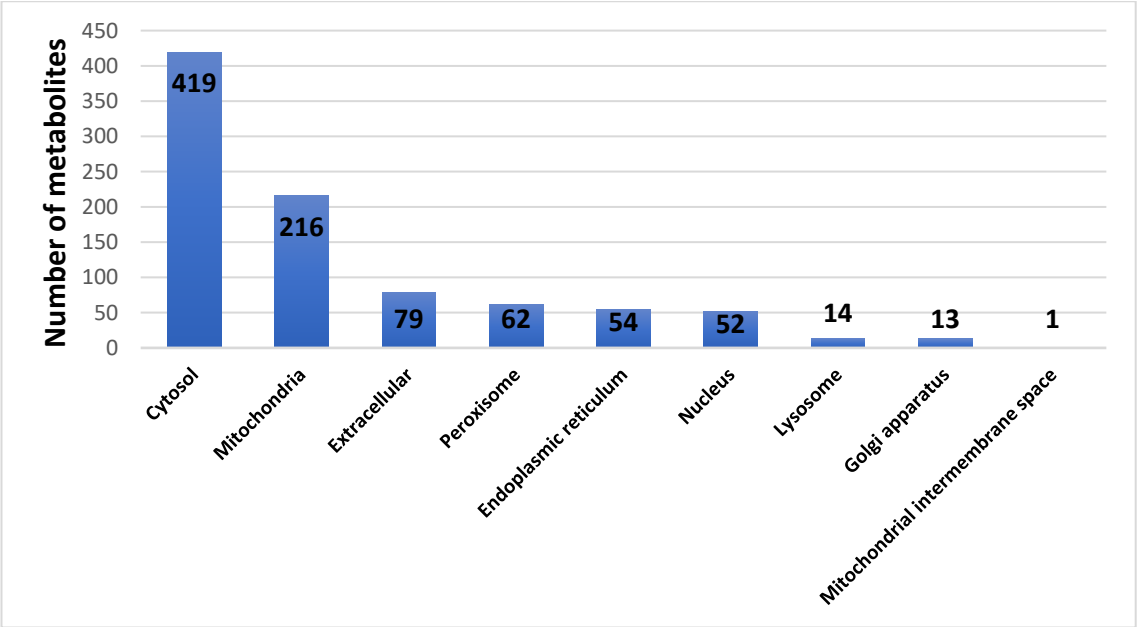


Figure 4.1 Compartmentalisation of metabolites in the skeletal muscle model
This bar chart shows the cellular compartments in the skeletal muscle model sorted by the number of metabolites in each compartment.

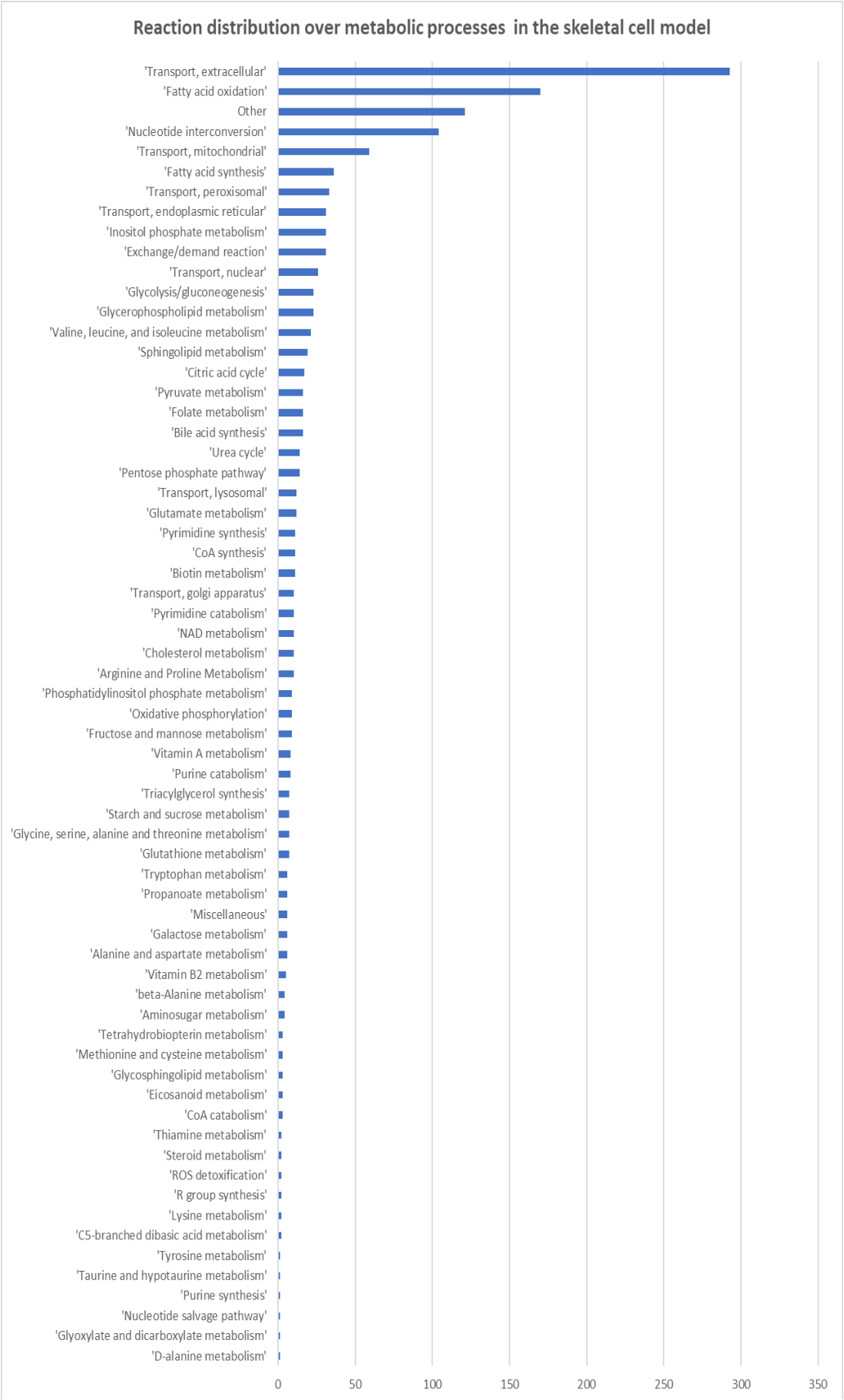


Figure 4.2 Reaction distribution over metabolic processes in the skeletal cell model
This bar chart shows the metabolic processes in the skeletal muscle model sorted by the number of reactions in each metabolic process.

4.3.2 Hypoxia and protein synthesis

Initially, the ATP production was assessed in normoxia, hypoxia and anoxia. When the oxygen uptake rate reached 10% of its normal value, the maximal ATP production rate was reduced by 88%, while in anoxia, the maximal ATP production rate dropped by 97% (Fig. 4.3). The small flux for maximal ATP production in anoxia was maintained entirely by glycolysis. Next, the model was optimised for protein synthesis. When oxygen availability was reduced to 10% of its normal value, the protein synthesis carried on at the normal rate. Only when the oxygen uptake rate approached 4.2% of the baseline value, the protein synthesis started to decrease until it reached 38% of its normal value in anoxia. The results showed that a severe reduction in oxygen uptake clearly reduced the protein synthesis rate in the skeletal muscle model (Fig. 4.4). Next, metabolic pathways leading to this reduction in protein synthesis rate were investigated. As the protein synthesis rate only started to drop at 4.2% oxygen, the analysis was performed with an oxygen uptake rate of 2% of its normal value.

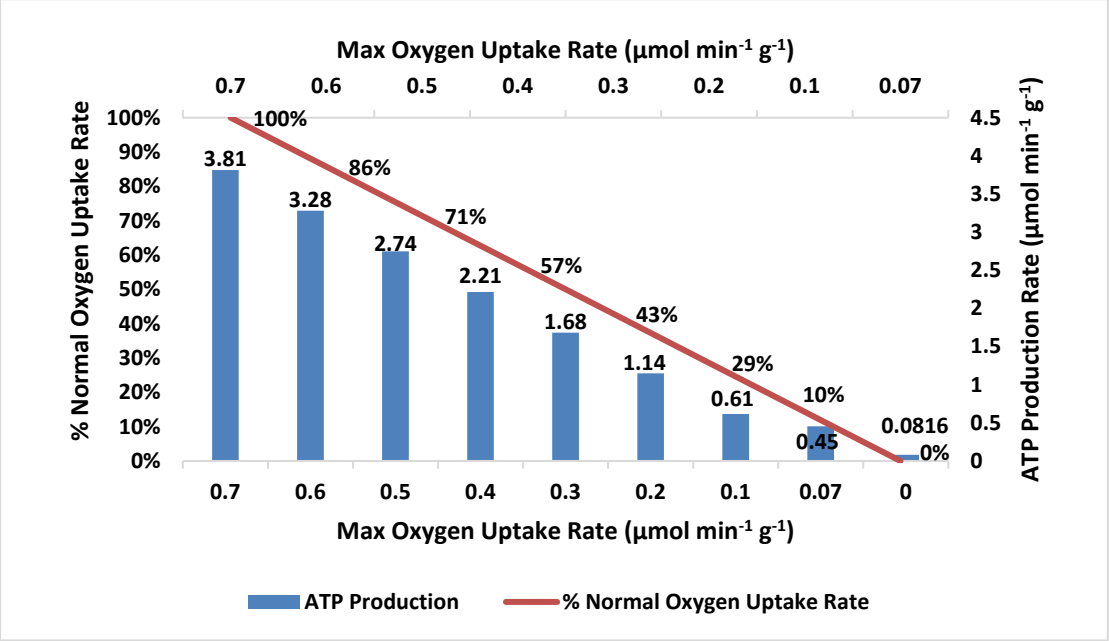


Figure 4.3 The relationship between ATP production and oxygen uptake in the skeletal muscle model
FBA was employed to calculate the maximal flux rate for protein synthesis while the oxygen uptake was progressively dropped from normoxia (100% oxygen uptake) to anoxia (0% oxygen uptake).

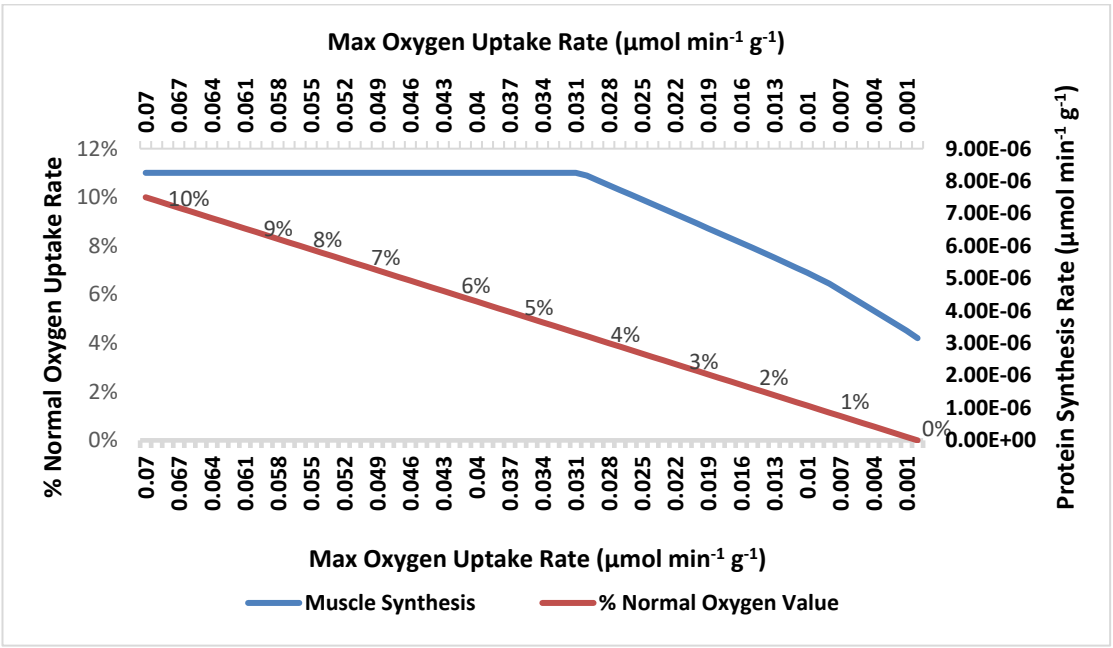


Figure 4.4 The relationship between protein synthesis and oxygen uptake in the skeletal muscle model
FBA was employed to calculate the maximal flux rate for protein synthesis while the oxygen uptake was progressively dropped.

Unsurprisingly, it was found that reactions associated with the electron transport chain (part of oxidative phosphorylation) had a lower flux in hypoxic conditions. As the model lies in the steady state and TCA cycle provides NADH to oxidative phosphorylation, the fluxes in the cycle were reduced too. It is important that both under hypoxia and normoxia, the carbon skeleton of valine is degraded by several pathways to produce succinyl-CoA which is further used in the TCA cycle. Under normoxia, the oxidation of valine is reduced and in consequence, the amount of succinyl-CoA which enters in the TCA cycle is also diminished. It is worth mentioning that under normoxia, the lower bound for valine uptake was hitting their limits as the valine influx was used for both valine degradation and protein synthesis equation. However, in hypoxia, the valine oxidation is reduced hence the reduction in valine influx. Lastly, the glucose uptake and fluxes through glycolytic reactions were elevated to maintain the amount of ATP which is needed by the protein synthesis reaction (Table 4.1).

Table 4.1 Flux rates during normoxia and hypoxia (2% oxygen) in the skeletal muscle model when optimising for myofibrillar protein synthesis

Reaction	Flux rate (µmol min^{-1} g^{-1} muscle)	
	Normoxia	Hypoxia
Glucose Uptake	0.0204	0.0425
HK	0.0204	0.0425
GPI	0.0204	0.0425
PFKM	0.0204	0.0425
ALDOA	0.0204	0.0425
TPI	0.0204	0.0425
GAPDH	0.0408	0.0850
PGK	0.0408	0.0850
PGAM	0.0408	0.0850
ENO	0.0408	0.0850
PKM	0.0408	0.0850
Lactate Efflux	0.0325	0.0816
PDH	0.0117	0.0058
LDHA	0.0441	0.0874
LDHB	0.0117	0.0058
CS	0.0095	0.0058
ACO2	0.0095	0.0058
IDH3	0.0045	0.0019
IDH2	0.0050	0.0039
OGDH	0.0061	0.0032
SUCLA1	0.0105	0.0058
SDHA	0.0105	0.0058
FH	0.0105	0.0058
MDH2	0.0140	0.0084
UQCR10	0.0770	0.0417
NDUFA10	0.0701	0.0313
ATP5F1B	0.1748	0.0939
COX	0.0385	0.0208

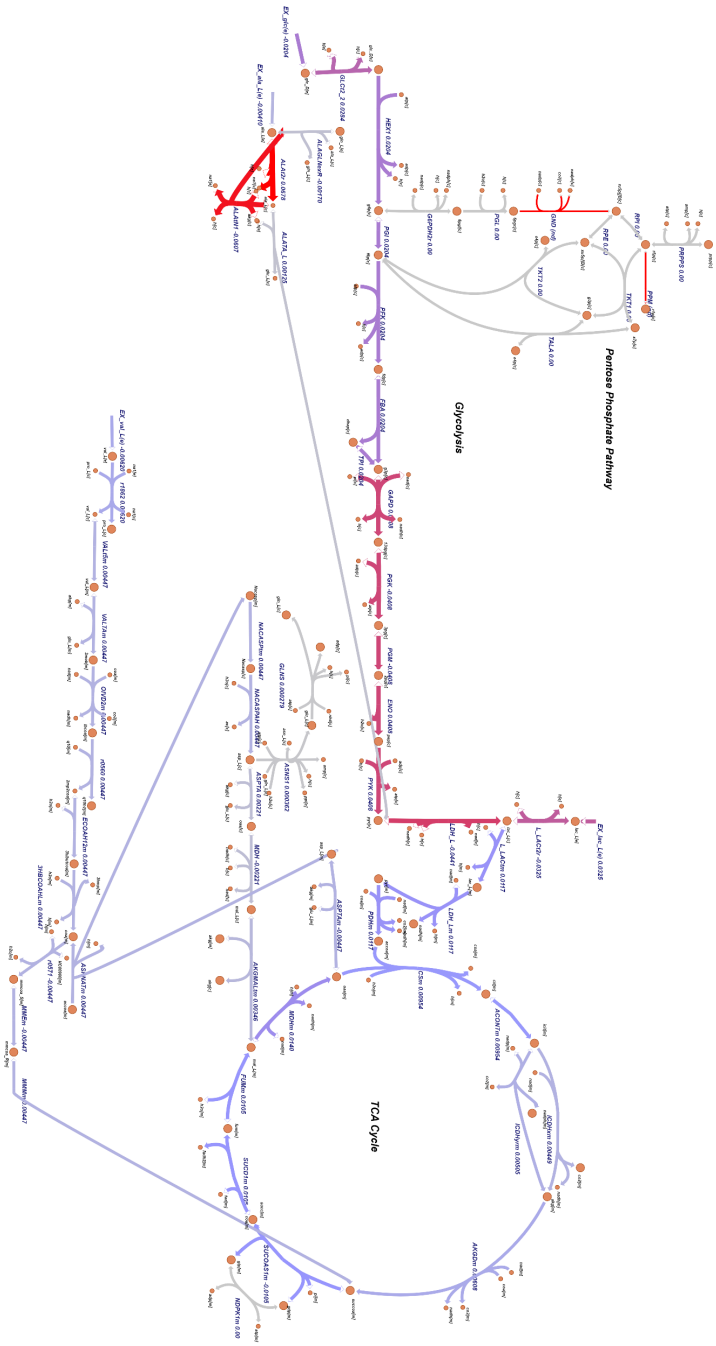
**The maximal flux rates for glycolytic fluxes (yellow), lactate processing fluxes (blue), TCA cycle (green) and electron transport chain (red) were derived using Flux Balance Analysis.*

When the protein synthesis objective function is maximised, due to certain proportions of each AA in the equation, some of these AAs will become rate-limiting. This is the case for alanine which becomes a limiting AA for protein synthesis. Hence with not enough import for alanine in the model, glutamine, which is used to produce alanine has its corresponding exchange reactions hitting its limits. The import of glutamine is used entirely to produce alanine. However, some glutamine is also used in the protein synthesis reaction hence it needs to be synthesised internally. Thus, oxaloacetate is used in aspartate transaminase to produce mitochondrial aspartate which is transported to cytosol. The cytosolic aspartate is then used to produce glutamate. Some of glutamate will go in the protein synthesis reaction, while the remaining is used to produce glutamine. Asparagine synthetase (glutamine hydrolysing) uses glutamine, ATP, water and aspartate to produce glutamate. Glutamate is then used in the protein synthesis reaction. Furthermore, in hypoxia, the TCA cycle and malate-aspartate shuttle are considerably reduced, and this leads to a reduced production of aspartate. With insufficient ATP to maintain the flux, the production of glutamate is also reduced. It is important to point out that under hypoxia, no glutamine is used to produce alanine and the entire amount of glutamine which is imported through the exchange reaction is used directly in the protein synthesis reaction (Fig. 4.5 a-b).

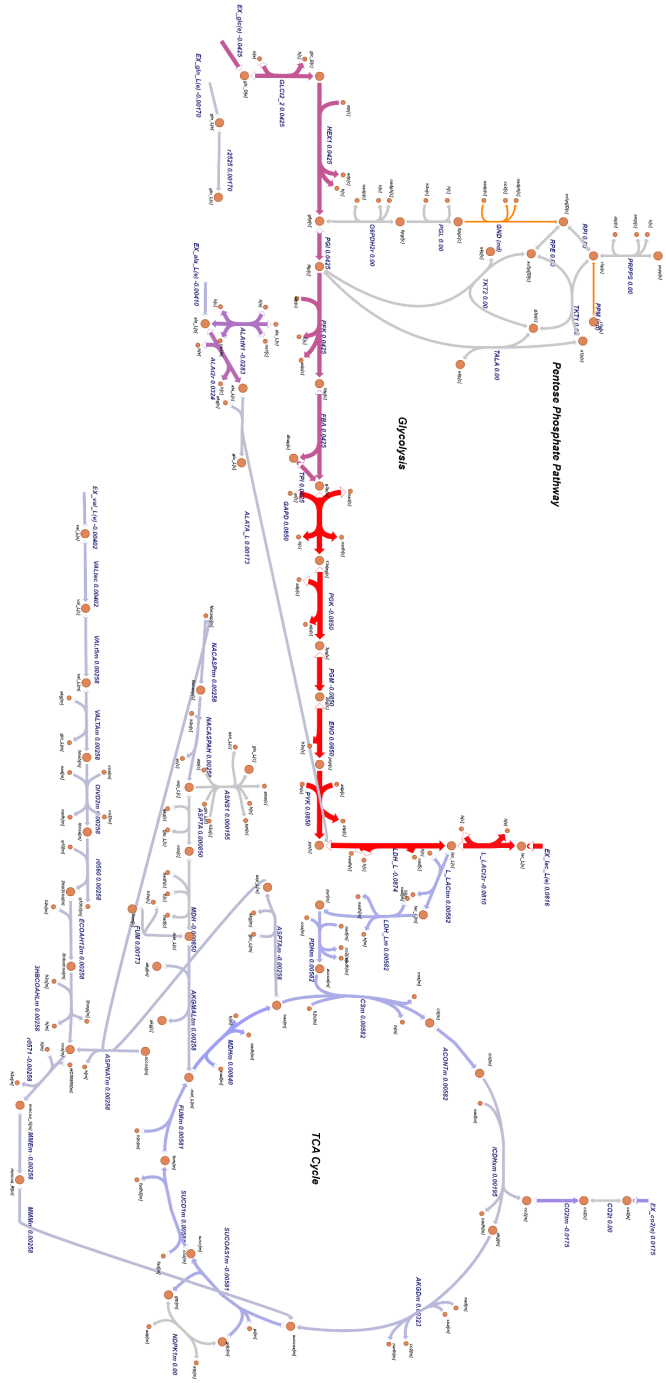
Furthermore, in normoxia, with enough alanine to support maximal protein synthesis, alanine transaminase uses alanine and α-ketoglutarate to produce glutamate and pyruvate. Pyruvate is then used in the PDH reaction to replenish the TCA cycle with acetyl-CoA. However, in hypoxia, ATP production and flux through the TCA cycle are reduced hence the internal production pathways for alanine are also affected. Most of the alanine is prioritised for protein synthesis and the flux for alanine transaminase is

reduced. However, pyruvate accumulates from the elevated glycolysis. This amount of pyruvate needs to be degraded or used by the model. Lactate dehydrogenase consists of two subunits: LDHA and LDHB. The first one has a higher affinity for pyruvate, converting pyruvate to lactate and NADH to NAD^+ in anaerobic conditions, while the second one has a higher affinity for lactate converting lactate to pyruvate and NADH to NAD^+ in aerobic conditions. Under normoxia, the entire amount of cytosolic pyruvate is converted to cytosolic lactate, which is further transported to mitochondria and converted to mitochondrial pyruvate by LDHB. Then, the mitochondrial pyruvate is converted to acetyl-CoA by PDH. In this scenario, there is no need for lactate to be secreted out of the model as the entire amount of pyruvate is converted to lactate and acetyl-CoA. LDHA was considerably elevated in hypoxic conditions, and the exchange reaction for lactate (lactate secretion) starts to pick up as oxygen level drops (Fig. 4.6).

LDHA also allows the removal of NADH that was not oxidised due to the limited capacity of oxidative phosphorylation. Thus, most of the pyruvate is converted to lactate, the NADH is oxidised, lactate is secreted through lactate exchange and PDH is reduced. However, the lactate secretion exchange reaction is constrained (upper bound) and the amount of pyruvate exceeds the upper bound. Thus, LDHB will try to convert the remaining lactate to pyruvate which is further converted to acetyl-CoA which refuels TCA cycle through PDH (Fig. 4.5 a-b). Under hypoxia, this shuttle between LDHA and LDHB clearly favours LDHA (higher flux) and PDH flux is reduced when compared to normoxia which is similar to that which happens *in vivo*.



(a)



(b)

Figure 4.5 The metabolic maps describing the TCA cycle and synthesis of rate-limiting AA when protein synthesis is optimised during normoxia (a) and hypoxia (b)
Under hypoxia, the glycolysis is elevated to maintain the ATP levels. Furthermore, most of the pyruvate is converted to lactate via LDHA and transported to the extracellular compartment to maintain the redox potential. A reduction in the TCA cycle and malate aspartate shuttle leads to a lower synthesis of alanine, glutamine and glutamate which are the rate-limiting AAs protein synthesis equation. Line thickness and colour (thin grey to bold red) are indicators of the flux values (low to high). The reactions which have no data in the model are coloured in orange. Escher (King et al. 2015) was used to generate the metabolic maps and visualise the flux distributions.

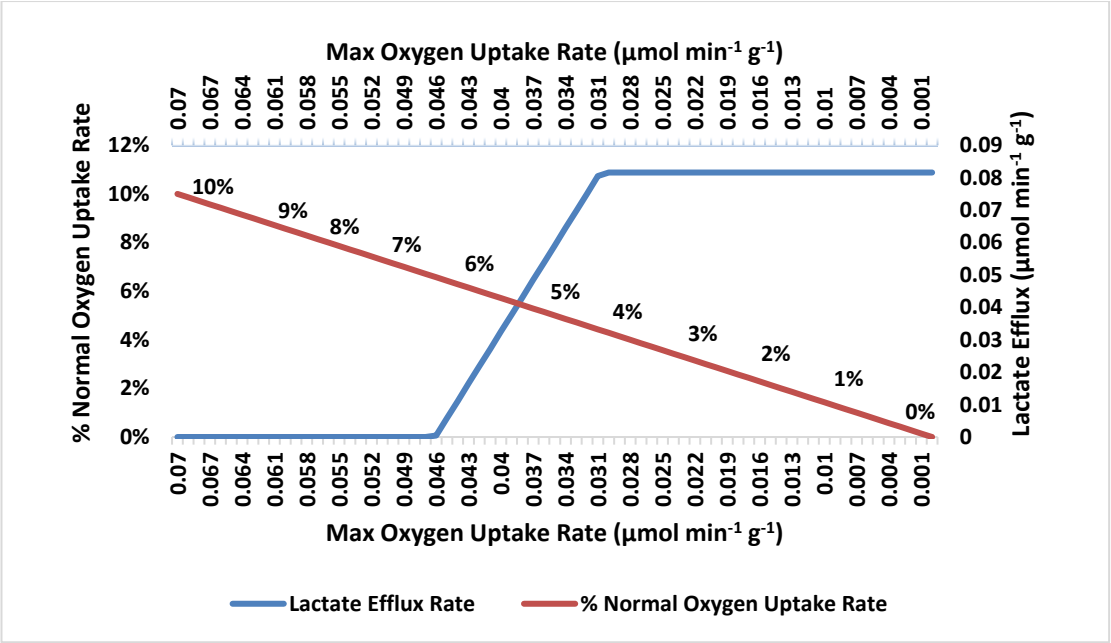


Figure 4.6 The relationship between lactate secretion and oxygen uptake in the skeletal muscle model
FBA was employed to calculate the maximal flux rate for lactate secretion while the oxygen uptake was progressively dropped.

Usually, under hypoxia, alternative carbon sources such as those provided by AAs can be used to produce building blocks which normally fuel the TCA cycle. Glutamine is converted to glutamate, which is in turn used to produce α-ketoglutarate that fuels the TCA cycle (glutaminolysis). Hence transaminase reactions usually work in the opposite way under hypoxia, using glutamate or aspartate to synthesise α-ketoglutarate. With alanine, glutamine and glutamate acting as limiting factors for the protein synthesis reaction and under hypoxia, the synthesis of these non-essential AAs is prioritised over the production of α-ketoglutarate. Thus, the recruited reactions from malate-aspartate shuttle such as aspartate transaminase are used to produce aspartate which is further used to synthesise glutamine and glutamate. Results showed that the oxygen reduction impacts protein synthesis, leading to an impaired TCA cycle which reduces the amount of aspartate, glutamine, glutamate and alanine.

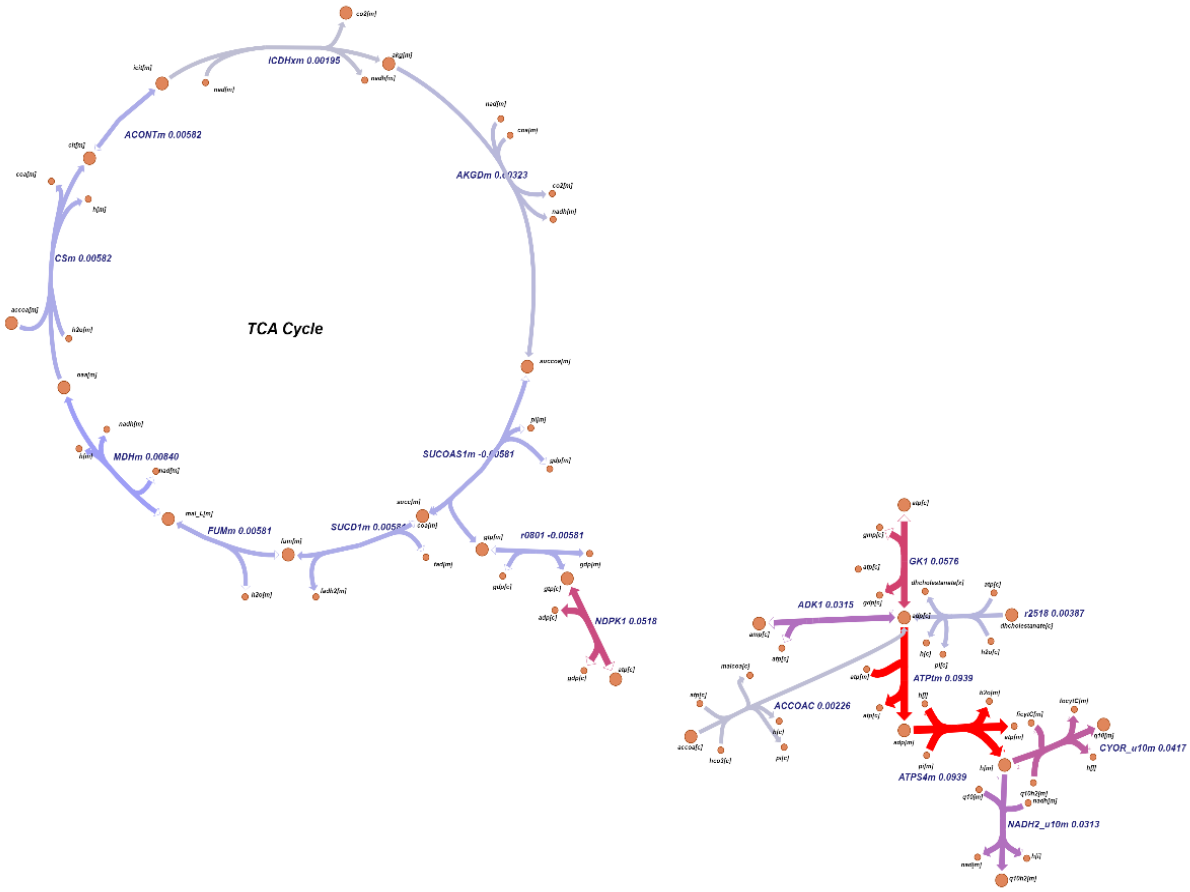
Flux spans revealed the model's reliance on glycolysis to produce ATP in low oxygen conditions (Tables 4.2-4.3). Under normoxic conditions, the model prioritises the reactions from the mitochondrial AA metabolism. By contrast, under hypoxic conditions, glycolytic reactions such as PGK, ENO, GAPDH, PGAM or TPI were ranked as the top reactions by flux span. These reactions were considered essential to hypoxic performance of the cell. Interestingly, guanylate kinase 1 (GK1) also showed up in the top reactions with a small relative flux span when optimising for myofibrillar protein synthesis. This reaction uses ATP to produce GDP and ADP. Next, the ADP is transported into the mitochondria where it is used in the ATP synthase (ATP5F1B) reaction. The ATP synthase reaction allows the hydrogen protons to flow into the mitochondrial matrix. These protons are then pumped out of the mitochondrial matrix in the NADH dehydrogenase (NDUFA10) and ubiquinol-6 cytochrome *c* reductase (UQCR10) reactions within oxidative phosphorylation. At the same time, the GTP (mitochondrial) is produced through SUCLG1 and transported into the cytosol where it is used in the protein synthesis reaction. It is important to point out that before its transportation to mitochondria, the ADP is synthesised in a few other reactions. In normoxia, these reactions include the nucleoside diphosphate kinase 1 (NME). However, under hypoxia, with less GTP derived from the TCA cycle, NME is used to produce GTP instead of ADP. With ATP production being reduced in hypoxia, FBA showed that GK1 had a lower flux compared to normoxia. However, with fewer alternative resources available, the ADP production through GK1 becomes even more essential for the network in hypoxia as it maintains the flux through NDUFA10 and UQCR10 reactions. Fig. 4.7 a-b describes these processes in greater detail.

Table 4.2 Reactions with smallest relative flux span when optimising skeletal muscle model for myofibrillar protein synthesis (Acto-Myosin complex) in normoxia

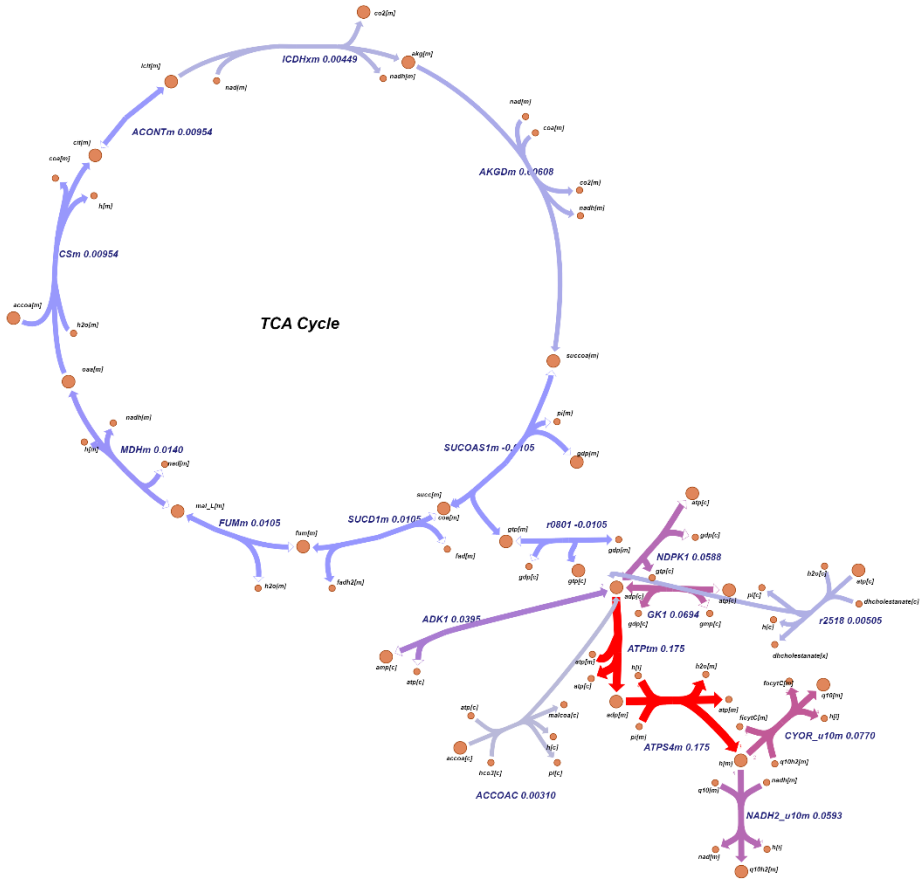
Rank	Rxn	Flux Span Value	Metabolic Pathway
1	BCKDHA	0.08	Valine, Leucine, and Isoleucine Metabolism
2	BCAT	0.08	Valine, Leucine, and Isoleucine Metabolism
3	HIBCH	0.08	Valine, Leucine, and Isoleucine Metabolism
4	HADH	0.08	Valine, Leucine, and Isoleucine Metabolism
5	HIBADH	0.08	Valine, Leucine, and Isoleucine Metabolism
6	MCEE	0.08	Valine, Leucine, and Isoleucine Metabolism
7	MMUT	0.08	Valine, Leucine, and Isoleucine Metabolism
8	ALT	0.31	Glutamate metabolism
9	ATP synthase	0.17	Oxidative Phosphorylation
10	GK1	0.43	Nucleotides
11	UQCR10	0.43	Oxidative Phosphorylation
12	COX	1.03	Oxidative Phosphorylation
13	NDUFA10	1.04	Oxidative Phosphorylation
14	ASNS (glutamine-hydrolysing)	1.09	Alanine and Aspartate Metabolism

Table 4.3 Reactions with smallest relative flux span when optimising skeletal muscle model for myofibrillar protein synthesis (Acto-Myosin complex) in hypoxia

Rank	Rxn	Flux Span Value	Metabolic Pathway
1	GK1	0.22	Nucleotides
3	PGK	0.33	Glycolysis
4	ENO	0.33	Glycolysis
5	GAPDH	0.33	Glycolysis
6	PGAM	0.33	Glycolysis
7	UQCR10	0.34	Oxidative Phosphorylation
8	TPI	0.37	Glycolysis
9	AK1	0.82	Nucleotides
10	NDUFA10	0.89	Oxidative Phosphorylation
11	COX	1.00	Oxidative Phosphorylation
12	HIBCH	1.47	Valine, Leucine, and Isoleucine Metabolism
13	HADH	1.47	Valine, Leucine, and Isoleucine Metabolism
14	HIBADH	1.47	Valine, Leucine, and Isoleucine Metabolism


TCA Cycle
ICDHxm 0.00195
ACONTm 0.00582
AKGDm 0.00323
CSm 0.00582
SUCOAS1m -0.00581
MDHm 0.00840
FUMm 0.00581
SUCD1m 0.00581
r0801 -0.00581
NDPK1 0.0518
GK1 0.0576
ADK1 0.0315
r2518 0.00387
ATPtm 0.0939
ACCOAC 0.00226
ATPS4m 0.0939
CYOR_u10m 0.0417
NADH2_u10m 0.0313

(a)



(b)

Figure 4.7 The metabolic map describing the role of guanylate kinase 1 in the skeletal muscle model under normoxia (a) and hypoxia (b)
GK1 is used to produce ADP. Under hypoxia, the GTP synthesis is prioritised. With fewer resources available to produce ADP and in hypoxia, GK1 becomes essential for the maintenance of the NDUFA10 and UQCR10 reactions within oxidative phosphorylation.

4.3.3 AA supplementation

Firstly, the results showed that in the single supplementation with EAA, lysine was the only EAA that slightly increased protein synthesis rate (Fig. 4.8). Lysine becomes the first EAA to hit its uptake limits. This is happening for two reasons: i) out of all EAA, lysine is the first limiting AA in the protein synthesis equation (Table 4.4); ii) lysine, is not synthesised internally by the model and it is imported directly in the objective function (Fig. 4.10). The lysine exchange reaction was subjected to robustness analysis to investigate the effects of flux alterations on the protein synthesis reaction. As the lysine AA is not synthesised internally, the results showed that protein synthesis is entirely dependent on the lysine influx and no intake of lysine would lead to zero protein synthesis (Fig. 4.9). None of the BCAAs impacted protein synthesis rate in single supplementation.

Next, when a supplementation with the three BCAAs (leucine, isoleucine and valine complex) was performed, the maximal protein synthesis rate did not increase either. However, when it comes to combinations of EAAs, isoleucine, leucine and valine appeared in all combinations that increased protein synthesis rate in skeletal muscle. Table 4.6 details the combinations of 1-9 EAAs that augmented protein synthesis rate. Apart from lysine and BCAAs, threonine, methionine, histidine and phenylalanine proved effective too in increasing the rate of protein synthesis when supplemented in combinations. Furthermore, using multiple EAA supplementation, a hierarchy for the rate-limiting AAs was obtained. This hierarchy is further discussed in section 4.4.3. The metabolic fate of the other EAAs is similar to that of lysine: they are transported from the extracellular compartment into the cytosol and used in the protein synthesis equation.

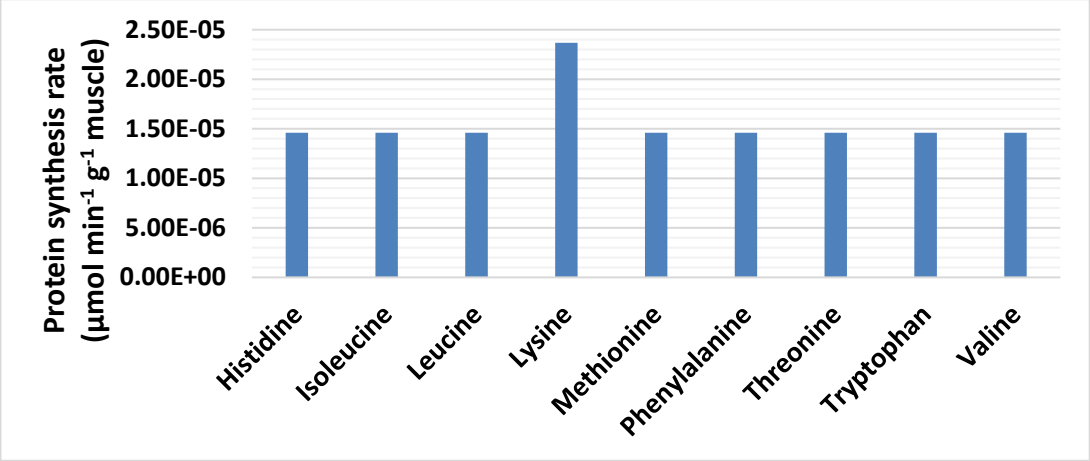


Figure 4.8 Single amino acid supplementation results in the skeletal muscle model
Each essential amino acid was supplemented individually and the maximal protein synthesis rate was derived using Flux Balance Analysis.

Table 4.4 The lower bounds in the skeletal muscle model vs. the actual uptake/secretion rates when optimising for protein synthesis

	Lb (max uptake rate mmol h^{-1} gDW^{-1})	Ub (max secretion rate mmol h^{-1} gDW^{-1})	Uptake Rate	Secretion Rate
Glc	-0.0916	1000	-0.0094	N/A
Gln	-1000	1000	-0.0125	N/A
Ala	-1000	1000	-0.0050	N/A
Arg	-1000	1000	-0.0031	N/A
Asn	-1000	1000	-0.0022	N/A
Gly	-1000	1000	-0.0035	N/A
His	-0.0040	1000	-0.0014	N/A
Leu	-0.0112	1000	-0.0051	N/A
Lys	-0.0047	1000	-0.0047	N/A
Ile	-0.0067	1000	-0.0041	N/A
Met	-0.0067	1000	-0.0023	N/A
Phe	-0.0164	1000	-0.050	N/A
Pro	-1000	1000	-0.0370	N/A
Ser	-1000	1000	-0.0034	N/A
Thr	-0.0085	1000	-0.0038	N/A
Trp	-0.0050	1000	4.81×10^{-4}	N/A
Tyr	-1000	1000	-0.0020	N/A
Val	-0.0062	1000	-0.0140	N/A
O_2	-0.728	1000	-0.0698	N/A

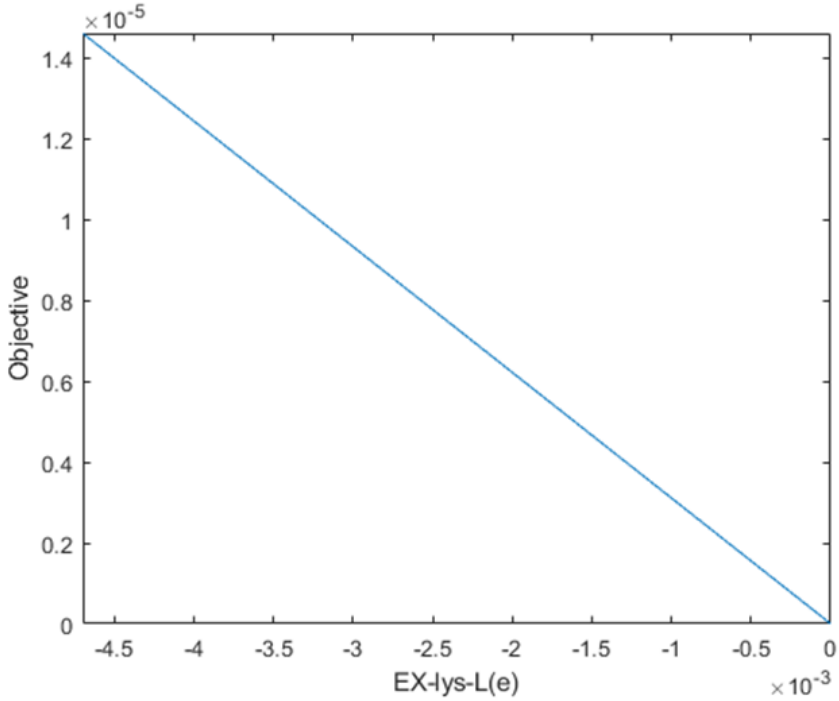


Figure 4.9 The objective function (protein synthesis reaction) as a function of the lysine exchange flux constraint
The lower bound for the lysine exchange constraint was varied from 0 to -4.7x10^{-3} and plotted against the corresponding FBA flux for protein synthesis.

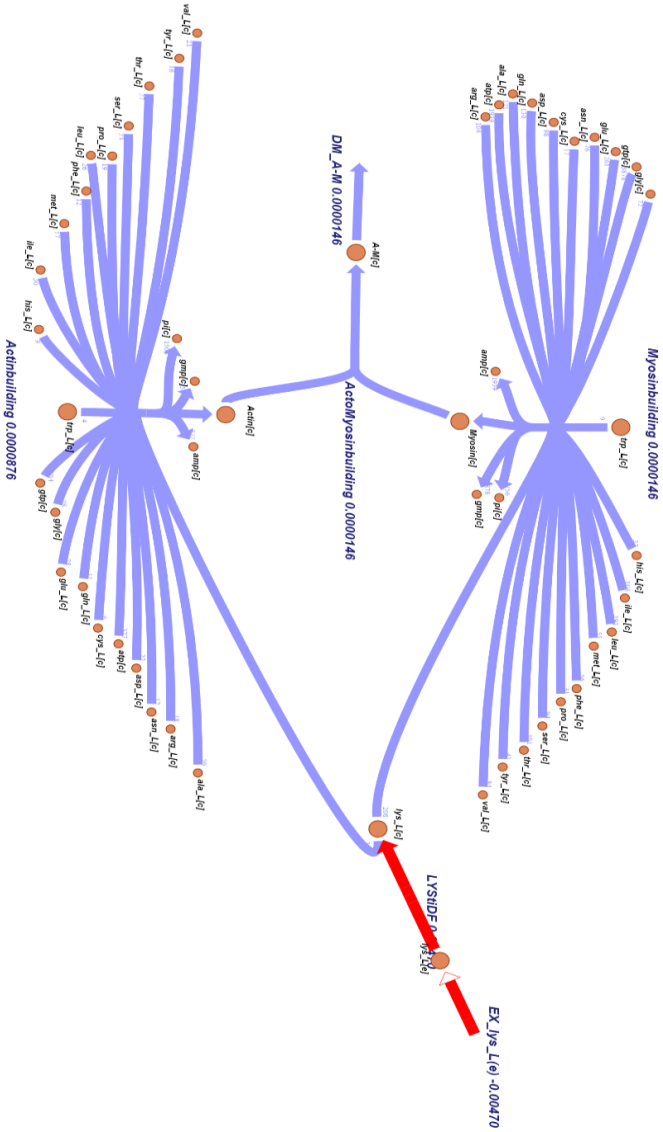


Figure 4.10 The metabolic map describing the mechanism by which lysine is used for protein synthesis
Lysine is imported from the extracellular compartment, transported into cytosol and directed to the protein synthesis equation.

Table 4.5 Multiple EAA supplementation in the skeletal muscle model

Number of EAAs	AA Combination	Protein synthesis rate (10^{-5} µmol min^{-1} g^{-1} muscle)
1	Lysine	2.37
2	Lysine, Isoleucine	2.95
3	Lysine, Isoleucine, Valine	3.22
4	Lysine, Isoleucine, Valine, Leucine	3.23
5	Lysine, Isoleucine, Valine, Leucine, Threonine	4.26
6	Lysine, Isoleucine, Valine, Leucine, Threonine, Methionine	4.39
7	Lysine, Isoleucine, Valine, Leucine, Threonine, Methionine, Histidine	12.8
8	Lysine, Isoleucine, Valine, Leucine, Threonine, Methionine, Histidine, Phenylalanine	15.1

**The essential amino acids were supplemented in all possible combinations to identify the most optimal combination that could increase protein synthesis rate. This table also ranks (ascending) the rate-limiting combinations based on their effect on protein synthesis rate.*

4.4 Discussion

Having divided the Results section (4.3) into three subsections, the Discussion section is also divided in this way.

4.4.1 Model generation

The skeletal muscle cell model was constructed using publicly available RNA-seq data. These data were obtained from an integrative "omics" study of the human proteome (Uhlen et al. 2015). In this study, samples of all major tissues and organs in the human body were used and RNA-seq was performed for 32 tissues. Data pertaining to skeletal muscle cell were downloaded and used to build the model as described in section 4.2.1. The current model captures the metabolism for the main macronutrients (AAs, carbohydrates, lipids) and micronutrients (vitamins) in the skeletal muscle cell. It contains 910 metabolites corresponding to 9 cell compartments and 1368 reactions distributed over 65 metabolic processes (Fig. 4.1- 4.2). Among these reactions, transport and extracellular reactions have the greatest proportion of total reactions (21.4%), followed by fatty acid oxidation (12.4% of total reactions). These numbers for the reaction distribution in skeletal muscle model are very similar to those associated with FB_{TGF} (myofibroblast). Whilst a direct comparison between the models of the two different cell types would be interesting, it is challenging because of the innate differences in their metabolic behaviour. However, it is interesting that both models predict that fat oxidation is the second largest contributor in terms of the number of reactions in the model. The large proportion of reactions involved in fatty acid oxidation might reflect the contractile phenotype of the two cells.

A CBM for skeletal muscle already exists (Nogiec et al. 2013). This model version is smaller in size and less developed from a metabolic point of view when compared to the

current model. The CBM proposed in 2013 has 341 metabolites, 374 reactions and it was aimed at capturing only the macronutrient metabolism. However, the contractile complexes for Type 1, 2a, 2x, and 2b myosin heavy chain isoforms were considered in the previous model, while the current model only contains Type 1 myosin isoforms. It is important to point out that myosin Type 2b is not expressed in human muscle (Bloemberg et al. 2012). Furthermore, the contractile unit contains troponin and tropomyosin together with actin and myosin, whereas the model proposed in this thesis contains only the two major proteins (actin and myosin). Thus, the previous model version is more developed with regards to the proteins which are part of the contractile unit. However, with the majority of contractile unit being comprised of actin and myosin, the production of Acto-Myosin complexes has successfully served as the main objective function (as described in section 2.2.7.3) for investigating the effects of hypoxia and AA supplementation on protein synthesis in the current skeletal muscle model.

4.4.2 Hypoxia

Protein synthesis is an energy dependent process and cells use energy in the form of ATP and GTP. Generally, low oxygen reduces the amount of ATP produced by the cell and protein synthesis drops. ATP can be synthesised either anaerobically through glycolysis to lactate, or aerobically through oxidative phosphorylation. Oxidative phosphorylation takes place in mitochondria and represents the principal source for substantial ATP production. Hypoxia blunts oxidative phosphorylation, and this leads to a greater reliance on (anaerobic) glycolysis to produce ATP. Furthermore, glucose as an anaerobic energy source is more oxygen efficient than fatty acids, hence a switch to glucose as an *anaerobic energy source* is another common feature of the hypoxic response. Most, if not all, organisms adapt to hypoxic conditions. However, those that

have evolved specifically to deal with low oxygen conditions may have cell-level adaptations. For instance, a study published in 1996 (Hochachka et al. 1996) reported that sea turtles are able to conserve oxygen by prioritising some vital ATP-consuming processes while reducing others, a feature not seen in rats. Furthermore, tumours, in which hypoxia is very common, are reported to find alternative pathways to produce ATP (Masoud et al. 2015). The discovery of hypoxia inducible factor HIF-1α (Semenza et al. 1994) provided further evidence for this metabolic switch from anaerobic to aerobic glycolysis. HIF-1α, a subunit of HIF-1 (hypoxia inducible factor 1), is a transcriptional factor that plays a crucial role in the organism adaptation to hypoxia. Under normoxic conditions, HIF-1α protein has a very short life of less than 5 minutes (Huang et al. 1998), but it accumulates quickly under low oxygen conditions (Salceda et al. 1997). It was reported that HIF-1α up-regulates several glycolytic enzymes during hypoxia (Kim et al. 2006, Solaini et al. 2010). These studies also reported that HIF-1α inhibits PDH enzyme which converts pyruvate to acetyl coenzyme A (acetyl-CoA). This is done by inducing the expression of PDK1, an isoform of pyruvate dehydrogenase kinase, which is a PDH inhibitor. Thus, acetyl-CoA is not allowed to enter in the TCA cycle while anaerobic glycolysis is elevated. In consequence, TCA cycle is supressed while pyruvate is converted to lactate via LDH.

The results from FBA showed that hypoxia significantly reduced ATP production. In anoxia, a small amount of ATP was produced through glycolysis. Similarly, the results from FBA showed that protein synthesis rate was considerably reduced in anoxia but not completely supressed as even with no oxygen the model will produce a small amount of ATP through glycolysis. GTP can be produced using ADP, and the direct import of AAs can be used to maintain some protein synthesis flux. Protein synthesis was not

reduced at 10% oxygen uptake and it started to gradually decrease only when oxygen uptake reached 4.2% of its normal value. It is also worth mentioning that representing hypoxia as a 10-20% reduction in the oxygen uptake rate is a common method in CBM (Edwards et al. 2014, Eyassu et al. 2017) but not entirely accurate. The relationship between oxygen concentration and oxygen consumption rate is closely related but the extent to which the rate of oxygen consumption decreases in response to progressive hypoxia is still debatable (Cobbs et al. 2018). It was reported that for some animals, the decline in oxygen consumption in relation to increased hypoxia is not linear (Cobbs et al. 2018). Moreover, the oxygen levels vary between tissues (McKeown 2014).

In hypoxia, due to low oxygen conditions, electron transport reactions and the TCA cycle were considerably reduced. Unsurprisingly, glycolytic fluxes were elevated as the model was trying to produce enough ATP to sustain protein synthesis. Glucose consumption also increased. *In vivo* studies have reported an increase in glucose uptake by 2-3 fold (Zierath et al. 1998) under hypoxic conditions. With lactate accumulating highly, the model converts most of the pyruvate to lactate which is secreted out of the model through lactate exchange reaction. As the amount of converted lactate exceeds the lactate secretion upper bound, the rest of cytosolic lactate is transported to mitochondria and then converted to mitochondrial pyruvate which is further converted to acetyl-CoA.

The elevation of LDHA under hypoxia is confirmed by the *in vivo* studies. Further, the flux rate for PDH is reduced in hypoxia compared to normoxia which is also partially consistent with the *in vivo* studies. However, HIF-1 is reported to inhibit the activity of PDH enzyme hence pyruvate is not converted to acetyl-CoA under hypoxia. Due to the limitations of the modelling work which is not able to capture the transcriptional

changes that are responsible for the inhibition of PDH by HIF-1, results only show a reduction in the flux rate for PDH.

The protein synthesis equation is limited by the availability of AAs in adequate ratios. Thus, certain AAs such as alanine become rate-limiting for protein synthesis equation. Alanine synthesis involves the use of other AAs which also become limiting factors in the protein synthesis reaction. In theory, non-essential AAs such as alanine should be produced internally by the model. The carbon skeletons for these AAs could be obtained from glycolysis or from intermediates in the TCA cycle. At the same time, pre-existing AAs can be broken down into an α-amino group and their carbon skeletons. Next, the α-amino group could be used to form new AAs through transaminase reactions. Although transaminase reactions are used to produce non-essential AAs, they are using the α-ketoglutarate or pyruvate fluxes, which, under hypoxia, are considerably reduced or prioritised for the replenishment of the TCA cycle. Hence the need for external provision of certain AAs through exchange reactions. Furthermore, the model lacks the reactions that should get rid of the remaining carbon skeletons which makes the synthesis of the non-essential AAs less possible.

A reduction in aerobic ATP production is considered the main driver for the reduction in protein synthesis under hypoxic conditions. However, in this case, the reduction is also attributed to an impaired synthesis of the building blocks of the protein synthesis equation owing to a reduction of aerobic ATP synthesis and TCA cycle. Under hypoxia, a trade-off between resources is employed by the model to support the production of the building blocks of the protein, TCA cycle and ATP production. Understanding this trade-off is challenging and the extent to which this trade-off plays a role biologically remains to be determined.

FVA and flux spans showed that, under hypoxic conditions, glycolytic fluxes outranked the electron transport chain reactions showing that in the absence of oxygen the model is forced to produce ATP through anaerobic glycolysis. Interestingly, GK1 also came up in the top reactions. In hypoxia, GTP production is prioritised over the synthesis of ADP which is further used to generate precursors for the reactions within oxidative phosphorylation. Hence the ADP synthesis via GK1 becomes essential for the network. With palmitate uptake being active, FBA also showed that, under hypoxic conditions, the fatty acids (palmitate) uptake is supressed. Fatty acids such as palmitate were used by the model in the fatty acid β-oxidation (FAO) which provides acetyl-CoA for TCA cycle and co-enzymes NADH and FADH2 that enter in the electron transport chain. As both the electron transport chain and TCA cycle are considerably reduced during hypoxia, both the fatty acid consumption and oxidation are ceased. It was reported that hypoxia reduces fatty acid oxidation (Morash et al. 2013) and fatty acid uptake (Musutova et al. 2018) in muscle cells. When fatty acid uptake is in place and under normoxic conditions, glucose uptake and glycolysis are supressed when optimising for myofibrillar protein synthesis. The relationship between fatty acids and glycolysis in muscle has been studied for many years. In a study based on rat muscle, published in 1963, the authors used the term “glucose and fatty acids cycle” and proposed a hypothesis that glucose and fatty acids “compete” for mitochondrial oxidation and this results in glucose uptake being reduced when fatty acids are present (Randle et al. 1963). Some more recent studies concluded that free fatty acids inhibit glucose transport and phosphorylation in human skeletal muscle (Krebs et al. 2001, Roden 2004). However, under hypoxia, glycolysis is once more utilised and fatty acid oxidation is ceased. This is in full agreement with studies confirming that during hypoxia the cell favours glucose uptake over fatty acid utilisation, and makes perfect sense from a modelling point of view: Under normoxic

conditions the model has sufficient oxygen for fatty acid oxidation, which increases the amount of acetyl-CoA, NADH and FADH2, precursors for TCA and oxidative phosphorylation. Thus, anaerobic glycolysis is not needed at all for the ATP production. When oxygen becomes limiting, the model can only run anaerobic glycolysis to produce ATP. It has been reported that under hypoxia, skeletal muscle cell favours glucose uptake over fatty acids utilisation (Horscroft et al. 2014).

Cells respond to hypoxia by reprogramming their metabolic pathways. The aim was to investigate these metabolic changes which impact protein synthesis rates in skeletal muscle. Although the model did not capture the transcriptional reprogramming of skeletal muscle during hypoxia, most of the metabolic pathways associated with this transcriptional reprogramming were clearly described by the simulations. The results showed that, in the absence of oxygen, the cell adapts its metabolism to produce the AAs in adequate proportions and synthesise ATP anaerobically through glycolysis. The model also prioritises the production of GTP over ADP synthesis to maximise protein synthesis.

In conclusion, the results showed that a considerable reduction in oxygen uptake causes a decline in protein synthesis rate which might provide the context for muscle wasting at high altitudes. However, it must be remembered that muscle mass is determined by the balance between protein synthesis and degradation and this model only focuses on the synthesis component. Furthermore, as it was outlined in the introductory chapters, beside a reduction in oxygen tension at high altitude, other factors such as reduction in appetite can influence the protein synthesis rate in muscle.

4.4.3 AA supplementation

It is still controversial whether supplementation with certain AAs can increase protein synthesis in muscle and if they can, which is the best combination that can trigger anabolic responses and increase protein synthesis rate. The model focused on single and multiple supplementation with EAAs to identify the relevant combinations. Non-essential AAs were left unconstrained as we were only interested in the EAA supplementation. A study published in 1998 showed that supplementation with non-essential AAs has no effect on the protein synthesis rate, while only supplementation with EAAs increased protein synthesis (Smith et al. 1998).

Single amino supplementation with EAAs showed that lysine was the only EAA that slightly augmented myofibrillar protein synthesis rate in skeletal muscle cell. A study reported that supplementation with combinations of EAAs including lysine increased protein synthesis rate in muscle (Tipton et al. 1999). It has also been reported that single supplementation with certain EAAs, including lysine, generated similar anabolic signalling pathways to valine and isoleucine (Atherton et al. 2010). In the current model, due to the proportion of lysine in the actin and myosin proteins and the lysine uptake rate derived from literature, the simulations suggested that supplementation with lysine alone can slightly increase the protein synthesis rate. These models rely on stoichiometry and the goal was to identify those EAAs which have the greatest weight on the protein synthesis equation based on their uptake rate which was derived from literature. Most probably, further experiments are needed to confirm the impact of lysine on protein turnover and muscle growth. When it comes to BCAAs, FBA showed no increase in the maximal protein synthesis rate when supplemented in combination of three (leucine, isoleucine and valine). Even though several reviews (Blomstrand et al.

2006, Norton et al. 2006) reported that BCAA supplementation leads to either increased protein synthesis or decreased protein degradation rate through activation of mTOR signalling pathways, there are also a number of studies that showed that supplementation with the BCAA alone has no anabolic effect (McNurlan et al. 1982, May et al. 1989, Louard et al. 1990). The model cannot capture the genetic control at transcriptional or translational level. Thus, it is impossible to assess the role of BCAAs in activating certain signalling pathways. However, when we performed multiple supplementation in combination, a supplement of lysine, leucine, isoleucine and valine increased protein synthesis rate. Thus, the model predicts that lysine is missing from this rate-limiting combination. It is uncertain whether supplementation with lysine and BCAAs is optimal for an increased protein synthesis rate and clearly further experimental work is needed to confirm this hypothesis. However, it was reported that supplementation with BCAA alone might not be enough to achieve maximum protein synthesis and other EAAs are needed in combination with BCAA to stimulate myofibrillar protein synthesis to near maximal levels (Jackman et al. 2017). Moreover, two studies reported that administration of BCAA alone not only was ineffective in stimulating the protein synthesis but also reduced it (Louard et al. 1990, Louard et al. 1995). With the earlier version of CBM for skeletal muscle (Nogiec et al. 2013) the authors performed similar simulations. In single AA supplementation, they reported only methionine to increase protein synthesis. Furthermore, they identified the combination of methionine, arginine, isoleucine, leucine, and phenylalanine, lysine and threonine (in this order) as the most optimal combination for increasing protein synthesis.

The current results also showed that BCAAs appeared in all combinations of EAAs (2-8) that increased protein synthesis rate. Threonine and methionine are also part of the

effective combinations of EAAs (5,6). The results also indicated that supplementation with a combination of all EAAs except tryptophan would reach the maximal protein synthesis rate. Although this result might sound trivial, a study that was previously mentioned reported that a supplementation with all essential AAs should be performed to trigger an accelerated anabolic response and generate a positive protein turnover (Wolfe 2017).

Apart from finding the “ideal” combination of AAs that can increase protein synthesis in skeletal muscle, another goal was to identify the limiting AAs. A rate-limiting AA is a particular EAA whose availability determines the maximal rate for protein synthesis. Supplementation with rate-limiting AAs is supposed to increase or maximise protein synthesis rate. In terms of the rate-limiting AAs, multiple supplementation with EAA provided us with a hierarchy: Lysine, BCAA (isoleucine, valine and leucine in this order), threonine, methionine, histidine, and phenylalanine. Each of the EAAs provided becomes a limiting term for the protein synthesis in this order. Based on the model simulations, lysine is the first limiting AA. Lysine is the first and most important limiting agent in horses and pigs and is reported to impact protein synthesis and muscle mass in these animals (Chang et al. 2005, Liao et al. 2015, Mastellar et al. 2016). Threonine and methionine were also reported to be limiting AAs in pigs (Wang et al. 2007). Even though in combinations with other EAAs it proved effective in stimulating protein turnover and muscle growth in humans, and is considered an essential AA, there is no clear experimental evidence to support that lysine is the first limiting AA in humans. However, it will be worth investigating how effective lysine is in combination with BCAAs and other EAAs.

It is important to point out that only EAAs were incorporated in this analysis. As discussed in the methods section, the non-essential AAs were made available for this analysis through their corresponding exchange reactions. Otherwise, non-essential AAs such as glutamine and alanine would also become rate-limiting AAs for protein synthesis before any other EAAs. However, this would happen due to certain reasons which were discussed in section 4.4.2. Hence the supplementation analysis would be biased by the incorporation of the non-essential AAs. This is because in this framework there are two types of essentiality for protein synthesis reaction: a "conditional" essentiality for the non-essential AAs which is determined by an overall lack of resources (degradation pathways, α-amino group etc.) and another type of essentiality for EAAs due to lack of synthesis pathways. However, the second type of essentiality has an important biological relevance as EAAs are not synthesised in the body and supplementation with these AAs is thought to enhance protein synthesis rate in muscle.

The fact that certain EAAs may be rate-limiting for protein synthesis in the human muscle has been discussed for many years and despite many scientific studies focusing on this fact, results are still controversial. The current results suggested that BCAAs can achieve maximal protein synthesis only in combinations with other EAAs. This is in full agreement with studies which reported this hypothesis (Jackman et al. 2017). Furthermore, the model predicted certain combinations that can impact protein synthesis rate in skeletal muscle cell. Simulations also showed that, with the exception of lysine, which can slightly increase protein synthesis rate in single supplementation, all other combinations need BCAA to reach maximal protein synthesis rate. Due to controversy surrounding AA supplementation and their impact on muscle growth, I acknowledge that further *in vivo* experiments are required to investigate the model

predictions. However, these modelling approaches can pave the way for using systems biology and COBRA methods for studying the effects of AA supplementation on muscle mass and growth.

Chapter 5: General discussion

This brief chapter provides a general overview and discussion of the work presented in this thesis, limitations of genome-scale CBMs in general, future directions of this area of research and future work that can be done to improve the modelling work performed in this thesis.

5.1 Recapitulation

The work in this thesis was focused on developing and analysing three *in silico* genome-scale models for FB_N, FB_{TGF} and skeletal muscle cells. The global human reconstruction (Recon 2.2) and RNA-seq transcriptomics data were used to generate context-specific models. The gene-protein-reaction (GPR) associations in Recon 2.2 were used to identify active (core) reactions in the required biological context and FASTCORMICS algorithm, an adaptation of FASTCORE, was used to generate context-specific models. The models were also manually refined and lower and upper bounds were imposed to limit the size of the solution space and improve model contextualisation. Constraint-based methods were used to analyse the model outputs, run simulations and generate predictions. RNA-seq technology was used to generate gene expression data in human lung fibroblast. Two datasets were generated: normal (FB_N) and TGF-β stimulated fibroblasts (FB_{TGF}). The latter one is considered a suitable *in vitro* model for myofibroblast cells. Two separate genome-scale models were derived from these data sets. These models were analysed separately using constraint-based methods and then compared. A biomass objective function was formulated to simulate proliferation. Furthermore, a collagen production reaction was added to the FB_N and FB_{TGF} models. Collagen synthesis and biomass production were optimised using FBA. FVA and flux spans were used to identify the most important reactions involved in collagen production and GEA was used to find

essential /critical genes for collagen synthesis. In separate modelling, publicly available RNA-seq data on human skeletal muscle was used to build a context-specific model. Reactions involved in the muscle building process were added to the model and optimised using FBA. The building of the contractile complex (Acto-Myosin complex) served as the main objective function to simulate muscle protein synthesis. This model was analysed under hypoxic conditions and AA supplementation to assess protein synthesis rate.

FBA results showed that both collagen synthesis and biomass production (proliferation) rates are higher in the FB_{TGF} model than in the FB_N model. Further, glycolysis rates were higher for the FB_{TGF} model. Next, the work focused more on the FB_{TGF} model to investigate the effect of TGF-β on collagen synthesis. Among a few other genes, GEA identified PHGDG, PSAT, SHTM, PSPH as critical/essential genes for collagen production in the FB_{TGF} model. These genes are involved in glycine and serine metabolism, so the model has revealed a critical role for glucose-derived serine/glycine biosynthesis as collagen intermediates. Furthermore, DEA showed that the genes involved in the *de novo* serine-glycine pathway are up-regulated in FB_{TGF}. Next, in the skeletal muscle model, the simulations showed that hypoxia reduced protein synthesis rate and a metabolic reprogramming which promotes AA synthesis and glycolysis over mitochondrial metabolism. Results also suggested that supplementation with certain combinations of AAs, including BCAA, could increase protein synthesis rate and facilitate skeletal muscle growth.

The success of the modelling approach was confirmed by making comparisons with experimental data in the literature. These models can be used for further simulations

with special attention to drug discovery, nutritional strategies and *in silico* target validation.

5.2 Limitations of genome-scale CBMs

Despite many strengths and applications of CBMs at steady state, there are still a number of limitations associated with these models. The main limitation is the lack of kinetic data. Genome-scale CBMs are purely based on reaction stoichiometry and reversibility and the steady-state assumption prevents the representation of the dynamic, transient nature of metabolism. For instance, FBA, the principal tool used for analysing CBMs, can only provide a single optimal solution of flux distributions for a certain objective function (reaction). Unlike CBMs, kinetic models are based on ordinary differential equations (ODE) and incorporate the rates of change for metabolite concentration as well as enzyme kinetic parameters. CBMs do not contain any information regarding metabolite and enzyme concentrations. Furthermore, these models cannot capture the effect of substrate concentration on enzyme activity as they lack enzymatic parameters such as Michaelis constant (Km) or maximal velocity (v_{max}). Further, when GEA is used to determine genes that are essential to a certain objective function (reaction), there is no information available about the regulatory changes that are associated with the deletion of those genes. However, integrating these parameters simultaneously into the high-throughput data is a difficult task and more advanced modelling methods are needed to achieve this objective.

5.3 Future directions

Given the limitations mentioned in the previous section, there are several ways to improve traditional genome-scale CBMs and bridge the gap between dynamic models

and genome-scale CBMs at steady state. To some extent, this section follows the logical approach proposed by Martins Conde Pdo et al. (2016).

Future developments of CBMs are based on two directions: integration of kinetic and regulatory parameters with current CBM methods and coupling multi-cellular or multi-tissue CBMs, the ultimate goal being the construction of a multi-cellular human kinetic model (Martins Conde Pdo et al. 2016). Even though this goal is far from being accomplished mainly due to the differences between kinetic and genome-scale CBMs, some progress was achieved to bring these two approaches together. For instance, several tools were developed to improve traditional FBA. FBA-derived tools such as regulatory FBA or rFBA (Covert et al. 2001), dynamic FBA or dFBA (Mahadevan et al. 2002), integrated FBA or iFBA (Covert et al. 2008), steady-state regulatory flux balance analysis or SR-FBA (Shlomi et al. 2007) were designed to integrate transcriptional regulation parameters with the traditional FBA using Boolean rules. Furthermore, PROM (Chandrasekaran et al. 2010) is an algorithmic tool that uses conditional probabilities to define the relationships between genes and transcriptional factors. In this way the regulatory parameters are integrated with the metabolic network. In this framework, FBA is only used for optimisation purposes. In a study published in 2010 (Fleming et al. 2010), the authors proposed a method to incorporate mass conservation at steady state, energy conservation and enzyme kinetics constraints into a single system of equations and inequalities. Another approach proposed by Pozo et al. (Pozo et al. 2012) was developed to optimise multiple objective functions in non-linear kinetic models. This algorithmic tool can be used to identify the global minimum for a metabolite synthesis rate while determining a minimum increase in concentration change of intermediate metabolites. Another *in silico* approach developed by Andreozzi et al. (Andreozzi et al.

2016) reduces the uncertainty that is usually associated with kinetic parameters and identifies the enzymes that should be addressed in order to obtain a desired physiological state of the organism. In this section I could only a brief overview of a few of the methods that have been developed in the recent years to integrate kinetic and regulatory parameters with traditional CBM. Research papers proposed by Martins Conde Pdo et al. (Martins Conde Pdo et al. 2016) and Lewis et al. (Lewis et al. 2012) provide a more detailed description of further tools that might be used.

Another future direction of CBMs is the building of multi-cellular or multi-tissue models by combining different models. When it comes to combining multiple CBMs, one of the main difficulties is dealing with different identifiers for metabolites reactions or genes (Pfau et al. 2016). Although, Systems Biology Markup Language (SBML) (Hucka et al. 2003) was initially designed to facilitate the interrogation of multiple models in a standardised format and a lot of effort was put in finding a common language between different databases, it is still difficult to deal with model parameters which are derived from different databases (Pfau et al. 2011). However, CBMs describing the interaction of multiple organisms (Bordbar et al. 2010, Heinken et al. 2013) and multi-tissues within the same organisms (Lewis et al. 2010, Bordbar et al. 2011) were constructed. Another, and probably most prominent challenge of this area of research is building a whole-body metabolic model. Although progression to this goal faces a lot of challenges, some attempts of combining CBMs with whole-body physiologically-based pharmacokinetic (PBPK) models have been made. PBPK models are used in the pharmaceutical industry to describe the absorption, distribution, metabolism, and excretion (ADME) of drugs in the human body. For instance, in a review proposed in 2012 (Krauss et al. 2012) a liver

CBM (Gille et al. 2010) was integrated with a whole-body PBMK to investigate the effects of tissue metabolic processes at a system level.

To conclude, the future of genome-scale CBM looks promising and I expect that more advanced computational tools will be developed to bridge the gap between CBMs and kinetic models. Although the objective of building a whole-body kinetic model for human metabolism is far from being met, the direction of travel suggests that this will ultimately be achieved.

5.4 Future work

It is important to point out that further work could be done to further improve the work that was performed in this thesis. Given the ideas and concepts outlined in this chapter, one interesting approach would be deriving kinetic parameters and transcriptional regulatory constraints and integrate them with the already-built CBMs to further reduce the solution space in all models. From a biological point of view, an interesting approach for the FB_{TGF} CBM would be building a multi-cell model of lung fibrosis to include myofibroblast and lung epithelium metabolism. Then, the approach could be extended to intracellular signalling as in (Aurich et al. 2012). This would be useful for interrogating epithelial cell biology as this cell type is notoriously difficult to culture/study pre-clinically and may provide insights into other respiratory diseases (e.g. COPD). Moreover, further mechanisms/targets that perturb myofibroblast biology with potential to additionally impact non-lung fibrotic diseases (hepatic, renal and cardiac fibrosis) could be identified. The overall aim for all of these approaches would be predicting important metabolic pathways and candidate drug targets that could become useful in treating IPF. With regards to the skeletal cell CBM, the contractile complex (objective function) which is currently formed of only actin and myosin, could be further

developed by adding other muscle proteins such as troponin, tropomyosin, titin and nebulin. This could improve the completeness of the skeletal muscle model.

Appendices Section

6.1 Biomass composition

Table 6.1.1

Main Breakdown

Mass unit: µg

	Lipids	Protein	Sugar	DNA	RNA	Total
	74.45	249	44.82	6	5	379.27
% Total	**19.63**	**65.65**	**11.81**	**1.58**	**1.31**	

Protein

Galacci, Higuchi (1966)

Lipids

Macarak, De la Llera et al. (1979), Harzer and Kustermann-Kuhn (2001)

Sugar

Umapathysivam, Hopwood et al. (2005)

DNA

(Bordbar et al. 2010))

RNA

Table 6.1.2

Protein

	Mol % Content	MW (g/mol)	Mass	Mass % Content	mmol/gDW
Alanine	0.0552	89.05	4.91	0.0376	**0.2774**
Arginine	0.1230	175.11	21.53	0.1648	**0.6182**
Asparagine	0.0371	132.05	4.89	0.0375	**0.1864**
Aspartic Acid	0.0642	132.04	8.47	0.0648	**0.3226**
Cysteine	0.0122	121.02	1.47	0.0113	**0.0613**
Glutamate	0.0731	146.05	10.67	0.0811	**0.3674**
Glutamine	0.0483	146.07	7.05	0.0540	**0.2427**
Glycine	0.0516	75.03	3.87	0.0296	**0.2593**
Histidine	0.0459	155.07	7.11	0.0544	**0.2306**
Isoleucine	0.0321	131.09	4.20	0.0322	**0.1613**
Leucine	0.0635	131.09	8.32	0.0632	**0.3191**
Lysine	0.1060	147.11	15.59	0.1193	**0.5327**
Methionine	0.0170	149.05	2.53	0.0193	**0.0854**
Phenylalanine	0.0309	165.08	5.10	0.0390	**0.1553**
Proline	0.0364	115.06	4.18	0.0329	**0.1829**
Serine	0.0432	105.04	4.53	0.0347	**0.2171**
Threonine	0.0379	119.06	4.51	0.0345	**0.1904**
Tryptophan	0.0105	204.09	2.14	0.0164	**0.0527**
Tyrosine	0.0239	181.07	4.32	0.033	**0.1201**
Valine	0.0438	117.08	5.12	0.0392	**0.2201**

Mol % Content was derived from Mohberg et al. (Mohberg et al. 1963)

Table 6.1.3

DNA

	Mol % Content	MW (g/mol)	Mass	Mass % Content	mmol/gDW
dAMP	0.29	349.23	103.02	0.31	**0.0140**
dCMP	0.20	307.19	62.97	0.18	**0.0097**
dGMP	0.20	347.22	71.18	0.21	**0.0097**
dTMP	0.29	322.21	95.05	0.28	**0.0140**

Mol % Content was derived from Bordbar et al. (Bordbar et al. 2010)

Table 6.1.4

RNA

	Mol % Content	MW (g/mol)	Mass	Mass % Content	mmol/gDW
AMP	0.18	345.06	62.45	0.18	**0.0070**
CMP	0.27	321.05	88.93	0.26	**0.0107**
GMP	0.36	361.06	131.78	0.38	**0.0141**
UMP	0.17	322.04	57.00	0.16	**0.0068**

Mol % Content was derived from East (East 1968)

Table 6.1.5

Sugar

	Mass	MW (g/mol)	mmol/gDW
Glycogen	0.085	162	0.52

Source: (Bordbar et al. 2010)

Table 6.1.6

Lipids

Lipid Fractions	% (weight)	AVG MW Weight	mmol/gDW
Neutral Lipids			
Myristic	0.028	228.37	**0.0002**
Palmitic	0.234	256.42	**0.0017**
Palmitoleic	0.026	254.40	**0.0002**
Stearic	0.29	284.47	**0.0020**
Oleic	0.39	282.46	**0.0027**
Cholesterol	10.2	386.7	**0.0517**
Diglycerides	1.4	615.82	**0.0044**
Triglycerides	6.1	820.29	**0.0145**
Phospholipids			
	43	791.37	**0.1066**
PE	23.3	742.63	**0.0615**
Sph	7.4	752.37	**0.0193**
PS	6.9	825.38	**0.0164**
PI	3.8	865.91	**0.0086**
PG	8.9	758.46	**0.0230**
LBPA	0.8	430.57	**0.0036**

% (weight) was derived from Weinstein et al. (Weinstein et al. 1969)

Table 6.1.7

ATP Maintenance

	mmol ATP/mmol aa
Protein synthesis and processing	4.30
RNA synthesis and processing	0.4
DNA synthesis and processing	1.37
Source: (Bordbar et al. 2010)	
	mmol/gDW
total protein	5.80
total protein	0.18
total DNA	0.07
	mmol ATP/gDW
protein	**25.00124**
RNA	**0.072309**
DNA	**0.099972**
total	**25.17353**

6.2 Uptake/secretion constraints

Table 6.2.1: Uptake/Secretion Constraints in the FB_N Model

Exchange reaction	Lb (max uptake rate mmol h^{-1} gDW^{-1})	Ub (max secretion rate mmol h^{-1} gDW^{-1})
Glc	-0.5280	0
Gln	-0.0290	1000
Arg	-0.0160	1000
Asn	-0.0041	1000
Asp	-0.0063	1000
His	-0.0032	1000
Leu	-0.0190	1000
Lys	-0.0120	1000
Ile	-0.0130	1000
Met	-0.0049	1000
Phe	-0.0058	1000
Ser	-0.0370	1000
Thr	-0.0084	1000
Trp	-0.0015	1000
Tyr	-0.0046	1000
Val	-0.0140	1000
Lac	0	0.695
O_2	-0.178	1000

Table 6.2.2: Uptake/Secretion Constraints in the FB_{TGF} Model

Exchange reaction	Lb (max uptake rate mmol gDW^{-1} h^{-1})	Ub (max secretion rate mmol gDW^{-1} h^{-1})
Glc	-1.056	0
Gln	-0.058	1000
Arg	-0.032	1000
Asn	-0.0082	1000
Asp	-0.0126	1000
His	-0.0064	1000
Leu	-0.0380	1000
Lys	-0.0240	1000
Ile	-0.0260	1000
Met	-0.0098	1000
Phe	-0.0116	1000
Ser	-0.0740	1000
Thr	-0.0168	1000
Trp	-0.0030	1000
Tyr	-0.0920	1000
Val	-0.0280	1000
Lac	0	1.74
O_2	-0.3560	1000

Table 6.2.3: Uptake/Secretion Constraints in the Skeletal Muscle Model

Exchange reaction	Lb (max uptake rate mmol gDW^{-1} h^{-1})	Ub (max secretion rate mmol gDW^{-1} h^{-1})
Glc	-0.0916	1000
Gln	-0.0017	1000
Ala	-0.0041	1000
Arg	-0.0052	1000
Asn	-0.0008	1000
Asp	-0.0008	1000
His	-0.0040	1000
Leu	-0.0112	1000
Lys	-0.0047	1000
Ile	-0.0067	1000
Met	-0.0067	1000
Phe	-0.0164	1000
Pro	-0.0029	1000
Ser	-0.0076	1000
Thr	-0.0085	1000
Trp	-0.0050	1000
Tyr	-0.0132	1000
Val	-0.0062	1000
Lac	0	0.0816
O_2	-0.728	1000
Palmitate	0.0333	1000

References

A Jensen, P., K. A Lutz and J. Papin (2011). TIGER: "Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks." BMC Syst Biol 5: 147

Acuña, V., F. Chierichetti, V. Lacroix, A. Marchetti-Spaccamela, M.-F. Sagot and L. Stougie (2009). "Modes and cuts in metabolic networks: Complexity and algorithms." Biosystems 95(1): 51-60.

Agren, R., S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew and J. Nielsen (2012). "Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT." PLoS Comput Biol 8(5): e1002518.

Alberch, P. (1991). "From genes to phenotype: dynamical systems and evolvability." Genetica 84(1): 5-11.

Andreozzi, S., L. Miskovic and V. Hatzimanikatis (2016). "iSCHRUNK- *In silico* approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks." Metab Eng 33: 158-168.

Atherton, P. J. and K. Smith (2012). "Muscle protein synthesis in response to nutrition and exercise." J Physiol 590(Pt 5): 1049-1057.

Atherton, P. J., K. Smith, T. Etheridge, D. Rankin and M. J. Rennie (2010). "Distinct anabolic signalling responses to amino acids in C2C12 skeletal muscle cells." Amino Acids 38(5): 1533-1539.

Aurich, M. K. and I. Thiele (2012). "Contextualization procedure and modeling of monocyte specific TLR signaling." PLoS One 7(12): e49978.

Baños, G., P. M. Daniel, S. R. Moorhouse and O. E. Pratt (1973). "The movement of amino acids between blood and skeletal muscle in the rat." J Physiol 235(2): 459-475.

Barbul, A. (2008). "Proline precursors to sustain Mammalian collagen synthesis." J Nutr 138(10): 2021s-2024s.

Baum, J. and H. S. Duffy (2011). "Fibroblasts and myofibroblasts: what are we talking about?" J Cardiovasc Pharmacol 57(4): 376-379.

Bazgir, B. (2017). "Satellite cells contribution to exercise mediated muscle." Cell J 18(4): 473-484.

Becker, S. A., A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson and M. J. Herrgard (2007). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox." Nat Protoc 2(3): 727-738.

Becker, S. A. and B. Ø. Palsson (2008). "Context-specific metabolic networks are consistent with experiments." PLoS Comput Biol 4(5): e1000082.

Berg, J. M., J. L. Tymoczko and L. Stryer (2002). Biochemistry, W.H. Freeman.

Bernard, K., N. J. Logsdon, S. Ravi, N. Xie, B. P. Persons, S. Rangarajan, J. W. Zmijewski, K. Mitra, G. Liu, V. M. Darley-Usmar and V. J. Thannickal (2015). "Metabolic reprogramming is required for myofibroblast contractility and differentiation." J Biol Chem 290(42): 25427-25438.

Bertsimas, D. and J. N. Tsitsiklis (1997). Introduction to Linear Optimization, Athena Scientific.

Beste, D. J., T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. E. Bushell, P. Wheeler, S. Klamt, A. M. Kierzek and J. McFadden (2007). "GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism." Genome Biol 8(5): R89.

Blaauw, B. and C. Reggiani (2014). "The role of satellite cells in muscle hypertrophy." J Muscle Res Cell M 35(1): 3-10.

Bloemberg, D. and J. Quadrilatero (2012). "Rapid determination of myosin heavy chain expression in rat, mouse, and human skeletal muscle using multicolor immunofluorescence Analysis." PLoS One 7(4): e35273.

Blomstrand, E., J. Eliasson, H. K. R. Karlsson and R. Köhnke (2006). "Branched-chain amino acids activate key enzymes in protein synthesis after physical exercise." J Nutr 136(1): 269S-273S.

Bochaton-Piallat, M. L. (2016). "The myofibroblast in wound healing and fibrosis: answered and unanswered questions." F1000Res 5: F1000 Faculty Rev-752

Bordbar, A., A. M. Feist, R. Usaite-Black, J. Woodcock, B. Ø. Palsson and I. Famili (2011). "A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology." BMC Syst Biol 5: 180.

Bordbar, A., N. E. Lewis, J. Schellenberger, B. Ø. Palsson and N. Jamshidi (2010). "Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions." Mol Syst Biol 6: 422.

Bordbar, A., J. M. Monk, Z. A. King and B. Ø. Palsson (2014). "Constraint-based models predict metabolic and associated cellular functions." Nat Rev Genet 15: 107.

Bordel, S. (2018). "Constraint based modeling of metabolism allows finding metabolic cancer hallmarks and identifying personalized therapeutic windows." Oncotarget 9(28): 19716-19729.

Borodina, I., P. Krabben and J. Nielsen (2005). "Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism." Genome Res 15(6): 820-829.

Borthwick, L. A., T. A. Wynn and A. J. Fisher (2013). "Cytokine mediated tissue fibrosis." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1832(7): 1049-1060.

Boyer, R. F. (2001). Concepts in Biochemistry, Wiley.

Boyer, S. J. and F. D. Blume (1984). "Weight loss and changes in body composition at high altitude." J Appl Physiol Respir Environ Exerc Physiol 57(5): 1580-1585.

British Lung Foundation. "Idiopathic pulmonary fibrosis statistics" https://www.blf.org.uk/support-for-you/idiopathic-pulmonary-fibrosis-ipf/statistics

Brunk, E., S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich, A. Prlić, A. Sastry, A. D. Danielsdottir, A. Heinken, A. Noronha, P. W. Rose, S. K. Burley, R. M. T. Fleming, J. Nielsen, I. Thiele and B. O. Palsson (2018). "Recon3D enables a three-dimensional view of gene variation in human metabolism." Nat Biotechnol 36(3): 272-281.

Selvarajah, B., I. Azuelos, M. Platé, D. Guillotin, E. J. Forty, G. Contento, H. V. Woodcock, M. Redding, A. Taylor, G. Brunori, P. F. Durrenberger, R. Ronzoni, A. D. Blanchard, P. F. Mercer, D. Anastasiou and R. C. Chambers (2019). "mTORC1 amplifies the ATF4-dependent de novo serine-glycine pathway to supply glycine during TGF-β_1-induced collagen biosynthesis." Sci Signal 12(582).

C. El-Saleh, S., K. Warber and J. Potter (1986). The role of tropomyosin-troponin in the regulation of skeletal muscle contraction. J Muscle Res Cell Motil 7(5): 387-404.

Carraro, F., C. A. Stuart, W. H. Hartl, J. Rosenblatt and R. R. Wolfe (1990). "Effect of exercise and recovery on muscle protein synthesis in human subjects." Am J Physiol 259(4 Pt 1): E470-476.

Chandrasekaran, S. and N. D. Price (2010). "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*." Proc Natl Acad Sci U S A 107(41): 17845-17850.

Chang, Y.-M. and H.-W. Wei (2005). "The effects of dietary lysine deficiency on muscle protein turnover in postweanling pigs." Asian-Australas J Anim Sci 18(9): 1326-1335.

Chen, S. J., W. Yuan, Y. Mori, A. Levenson, M. Trojanowska and J. Varga (1999). "Stimulation of type I collagen transcription in human skin fibroblasts by TGF-β: involvement of Smad 3." J Invest Dermatol 112(1): 49-57.

Chicco, A. J., C. H. Le, E. Gnaiger, H. C. Dreyer, J. B. Muyskens, A. D'Alessandro, T. Nemkov, A. D. Hocker, J. E. Prenni, L. M. Wolfe, N. M. Sindt, A. T. Lovering, A. W. Subudhi and R. C. Roach (2018). "Adaptive remodeling of skeletal muscle energy metabolism in high-altitude hypoxia: Lessons from AltitudeOmics." J Biol Chem 293(18): 6659-6671.

Chitturi, R. T., A. M. Balasubramaniam, R. A. Parameswar, G. Kesavan, K. T. Haris and K. Mohideen (2015). "The role of myofibroblasts in wound healing, contraction and its clinical implications in cleft palate repair." J Int Oral Health 7(3): 75-80.

Clark, D. A. and R. Coker (1998). "Molecules in focus Transforming growth factor-beta (TGF-β)." Int J Biochem Cell Biol 30(3): 293-298.

Cobbs, G. A. and J. E. Alexander, Jr. (2018). "Assessment of oxygen consumption in response to progressive hypoxia." PLoS One 13(12): e0208836.

Consortium, T. U. (2018). "UniProt: a worldwide hub of protein knowledge." Nucleic Acids Res 47(D1): D506-D515.

Cooper, G. M. (2000). The Cell - A Molecular Approach 2nd Edition, Sunderland (MA): Sinauer Associates.

Covert, M. W., I. Famili and B. Ø. Palsson (2003). "Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology?" Biotechnol Bioeng 84(7): 763-772.

Covert, M. W. and B. Ø. Palsson (2003). "Constraints-based models: regulation of gene expression reduces the steady-state solution space." J Theor Biol 221(3): 309-325.

Covert, M. W., C. H. Schilling and B. Ø. Palsson (2001). "Regulation of gene expression in flux balance models of metabolism." J Theor Biol 213(1): 73-88.

Covert, M. W., N. Xiao, T. J. Chen and J. R. Karr (2008). "Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*." Bioinformatics 24(18): 2044-2050.

Crick, F. (1970). "Central dogma of molecular biology." Nat 227(5258): 561-563.

Dantzig, G. B. and S. U. S. O. Laboratory (1987). Origins of the Simplex Method, Defense Technical Information Center.

Darby, I., O. Skalli and G. Gabbiani (1990). "Alpha-smooth muscle actin is transiently expressed by myofibroblasts during experimental wound healing." Lab Invest 63(1): 21-29.

Darby, I. A. and T. D. Hewitson (2007). Fibroblast differentiation in wound healing and fibrosis. International Review of Cytology, Academic Press. 257: 143-179.

Darby, I. A., B. Laverdet, F. Bonté and A. Desmoulière (2014). "Fibroblasts and myofibroblasts in wound healing." Clin Cosmet Investig Dermatol 7: 301-311.

Dartt, D. A., J. Besharse, R. Dana, B. A. Battelle, D. Beebe, P. Bex, P. Bishop, D. Bok, P. D'Amore and H. Edelhauser (2010). Encyclopedia of the Eye, Elsevier Science.

de Boer, M. D., C. N. Maganaris, O. R. Seynnes, M. J. Rennie and M. V. Narici (2007). "Time course of muscular, neural and tendinous adaptations to 23 day unilateral lower-limb suspension in young men." J Physiol 583(Pt 3): 1079-1091.

De Smet, R. and K. Marchal (2010). "Advantages and limitations of current network inference methods." Nat Res Microbiol 8: 717.

Desmoulière, A., C. Chaponnier and G. Gabbiani (2005). "Perspective Article: Tissue repair, contraction, and the myofibroblast." Wound Repair and Regeneration 13(1): 7-12.

Desmoulière, A., A. Geinoz, F. Gabbiani and G. Gabbiani (1993). "Transforming growth factor-β 1 induces alpha-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts." J. Cell Biol 122(1): 103-111.

Desmoulière, A., M. Redard, I. Darby and G. Gabbiani (1995). "Apoptosis mediates the decrease in cellularity during the transition between granulation tissue and scar." Am J Pathol 146(1): 56-66.

Dietmair, S., M. P. Hodson, L. E. Quek, N. E. Timmins, P. Gray and L. K. Nielsen (2012). "A multi-omics analysis of recombinant protein production in Hek293 cells." PLoS One 7(8): e43394.

Dixon, K. and E. Holmes (1935). "Mechanism of the Pasteur Effect." Nat 135: 995.

Duarte, N. C., S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas and B. Ø. Palsson (2007). "Global reconstruction of the human metabolic network based on genomic and bibliomic data." Proc Natl Acad Sci U S A 104(6): 1777-1782.

Dugina, V., L. Fontao, C. Chaponnier, J. Vasiliev and G. Gabbiani (2001). "Focal adhesion features during myofibroblastic differentiation are controlled by intracellular and extracellular factors." J Cell Sci 114(18): 3285-3296.

Durinck, S., P. T. Spellman, E. Birney and W. Huber (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt." Nat Protoc 4(8): 1184-1191.

East, J. L. (1968). "Nucleotide composition of ribonucleic acid by spectral analysis and thin-layer chromatography." Analytical Biochemistry 24(3): 409-418.

Edwards, J. S., R. U. Ibarra and B. Ø. Palsson (2001). "*In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data." Nat Biotechnol 19: 125.

Edwards, J. S. and B. Ø. Palsson (1999). "Systems properties of the *Haemophilus influenzae* Rd metabolic genotype." J Biol Chem 274(25): 17410-17416.

Edwards, J. S. and B. Ø. Palsson (2000). "The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities." Proc Natl Acad Sci USA 97(10): 5528-5533.

Edwards, J. S. and B. Ø. Palsson (2000). "Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions." BMC Bioinformatics 1: 1.

Edwards, J. S. and B. Ø. Palsson (2000). "Robustness analysis of the *Escherichia coli* metabolic network." Biotechnol Prog 16(6): 927-939.

Edwards, L. M. (2017). "Metabolic systems biology: a brief primer." J Physiol 595(9): 2849-2855.

Edwards, L. M., M. I. Sigurdsson, P. A. Robbins, M. E. Weale, G. L. Cavalleri, H. E. Montgomery and I. Thiele (2014). "Genome-scale methods converge on key mitochondrial genes for the survival of human cardiomyocytes in hypoxia." Circulation: Cardiovascular Genetics 7(4): 407-415.

Etheridge, T., P. J. Atherton, D. Wilkinson, A. Selby, D. Rankin, N. Webborn, K. Smith and P. W. Watt (2011). "Effects of hypoxia on muscle protein synthesis and anabolic signaling at rest and in response to acute resistance exercise." Am J Physiol Endocrinol Metab 301(4): E697-702.

Eyassu, F. and C. Angione (2017). "Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism." Royal Society open science 4(10): 170360-170360.

Eyden, B. P. (1993). "Brief Review of the fibronexus and its significance for myofibroblastic differentiation and tumor diagnosis." Ultrastructural Pathology 17(6): 611-622.

Fajas, L. and A. Giralt (2018). "Metabolic adaptation to cell growth and proliferation in normal and pathological conditions. " Frontiers Media SA.

Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. Ø. Palsson (2007). "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol 3: 121.

Fernandez, I. E. and O. Eickelberg (2012). "The impact of TGF-β on lung fibrosis: from targeting to biomarkers." Proc Am Thorac Soc 9(3): 111-116.

Fischer, H. P. (2008). "Mathematical modeling of complex biological systems: from parts lists to understanding systems behavior." Alcohol Res Health 31(1): 49-59.

Fleming, R. M. T., I. Thiele, G. Provan and H. P. Nasheuer (2010). "Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism." J Theor Biol 264(3): 683-692.

Folger, O., L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin and T. Shlomi (2011). "Predicting selective drug targets in cancer through metabolic networks." Mol Syst Biol 7: 501.

Forster, J., I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen (2003). "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network." Genome Res 13(2): 244-253.

Friboulet, A. and D. Thomas (2005). "Systems Biology—an interdisciplinary approach." Biosens Bioelectron 20(12): 2404-2407.

Funato, N., K. Moriyama, Y. Baba and T. Kuroda (1999). "Evidence for apoptosis induction in myofibroblasts during palatal mucoperiosteal repair." J Dent Res 78(9): 1511-1517.

Gabbiani, G., G. B. Ryan and G. Majno (1971). "Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction." Experientia 27(5): 549-550.

Gailit, J., M. J. Marchese, R. R. Kew and B. L. Gruber (2001). "The differentiation and function of myofibroblasts is regulated by mast cell mediators." J Invest Dermatol 117(5): 1113-1119.

Gallaci, Higuchi (1966) United States Army Biological Center, Fort Detrick, MD.

Ganapathy-Kanniappan, S. and J. F. Geschwind (2013). "Tumor glycolysis as a target for cancer therapy: progress and prospects." Mol Cancer 12: 152.

Garlick, P. J. (2005). "The role of leucine in the regulation of protein metabolism." J Nutr 135(6): 1553S-1556S.

Gass, S. I. (2003). Linear Programming: Methods and Applications, Dover Publications.

Ge, H., A. J. Walhout and M. Vidal (2003). "Integrating 'omic' information: a bridge between genomics and systems biology." Trends Genet 19(10): 551-560.

Gille, C., C. Bölling, A. Hoppe, S. Bulik, S. Hoffmann, K. Hübner, A. Karlstädt, R. Ganeshan, M. König, K. Rother, M. Weidlich, J. Behre and H. G. Holzhütter (2010). "HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology." Mol Syst Biol 6: 411.

Gomes, A. V., J. D. Potter and D. Szczesna-Cordary (2002). "The role of troponins in muscle contraction." IUBMB Life 54(6): 323-333.

Gonzalez, A. C. O. (2016). "Wound healing - A literature review." 91(5): 614-620.

Guarente, L. (1993). "Synthetic enhancement in gene interaction: a genetic tool come of age." Trends Genet 9(10): 362-366.

Gudmundsson, S. and I. Thiele (2010). "Computationally efficient flux variability analysis." BMC Bioinformatics 11: 489.

Gueorguieva, I., A. L. Cleverly, A. Stauber, N. S. Pillay, J. A. Rodon, C. P. Miles, J. M. Yingling and M. M. Lahn (2014). "Defining a therapeutic window for the novel TGF-β inhibitor LY2157299 monohydrate based on a pharmacokinetic/pharmacodynamic model." Br J Clin Pharmacol 77(5): 796-807.

Hagen, J. B. (2000). "The origins of bioinformatics." Nat Rev Genet 1(3): 231-236.

Hamanaka, R. B. and N. S. Chandel (2012). "Targeting glucose metabolism for cancer therapy." J Exp Med 209(2): 211-215.

Hamernik, D. L. (2019). "Farm animals are important biomedical models." Animal Frontiers 9(3): 3-5.

Harris, T. W., V. Arnaboldi, S. Cain, J. Chan, W. J. Chen, J. Cho, P. Davis, S. Gao, C. A. Grove, R. Kishore, R. Y. N. Lee, H.-M. Muller, C. Nakamura, P. Nuin, M. Paulini, D. Raciti, F. H. Rodgers, M. Russell, G. Schindelman, K. V. Auken, Q. Wang, G. Williams, A. J. Wright, K. Yook, K. L. Howe, T. Schedl, L. Stein and P. W. Sternberg (2019). "WormBase: a modern model organism information resource." Nucleic Acids Res 48(D1): D762-D767.

Harzer, K. and B. Kustermann-Kuhn (2001). "Quantified increases of cholesterol, total lipid and globotriaosylceramide in filipin-positive Niemann-Pick type C fibroblasts." Clin Chim Acta 305(1-2): 65-73.

Heinken, A., S. Sahoo, R. M. Fleming and I. Thiele (2013). "Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut." Gut Microbes 4(1): 28-40.

Hinz, B. (2009). "Tissue stiffness, latent TGF-β1 Activation, and mechanical signal transduction: Implications for the pathogenesis and treatment of fibrosis." Current Rheumatology Reports 11(2): 120.

Hocevar, B. A., T. L. Brown and P. H. Howe (1999). "TGF-β induces fibronectin synthesis through a c-Jun N-terminal kinase-dependent, Smad4-independent pathway." Embo j 18(5): 1345-1356.

Hochachka, P. W., L. T. Buck, C. J. Doll and S. C. Land (1996). "Unifying theory of hypoxia tolerance: molecular/metabolic defense and rescue mechanisms for surviving oxygen lack." Proc Natl Acad Sci USA 93(18): 9493-9498.

Holm, L., M. L. Haslund, P. Robach, G. van Hall, J. A. Calbet, B. Saltin and C. Lundby (2010). "Skeletal muscle myofibrillar and sarcoplasmic protein synthesis rates are affected differently by altitude-induced hypoxia in native lowlanders." PLoS One 5(12): e15606.

Hoppeler, H., H. Howald and P. Cerretelli (1990). "Human muscle structure after exposure to extreme altitude." Experientia 46(11-12): 1185-1187.

Horscroft, J. A. and A. J. Murray (2014). "Skeletal muscle energy metabolism in environmental hypoxia: climbing towards consensus." Extreme Physiology & Medicine 3(1): 19.

Huang, L. E., J. Gu, M. Schau and H. F. Bunn (1998). "Regulation of hypoxia-inducible factor 1α is mediated by an O_2-dependent degradation domain via the ubiquitin-proteasome pathway." Natl Acad Sci USA 95(14): 7987-7992.

Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, Goryanin, II, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner and J. Wang (2003). "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models." Bioinformatics 19(4): 524-531.

Hunt, S. E., W. McLaren, L. Gil, A. Thormann, H. Schuilenburg, D. Sheppard, A. Parton, I. M. Armean, S. J. Trevanion, P. Flicek and F. Cunningham (2018). "Ensembl variation resources." Database (Oxford) 2018.

Hyduke, D. R., N. E. Lewis and B. Ø. Palsson (2013). "Analysis of omics data with genome-scale models of metabolism." Mol Biosyst 9(2): 167-174.

Ibarra, R. U., J. S. Edwards and B. Ø. Palsson (2002). "*Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth." Nat 420(6912): 186-189.

Ignotz, R. A. and J. Massague (1986). "Transforming growth factor-β stimulates the expression of fibronectin and collagen and their incorporation into the extracellular matrix." J Biol Chem 261(9): 4337-4345.

Illsley, M. C., J. H. Peacock, R. J. McAnulty and J. R. Yarnold (2000). "Increased collagen production in fibroblasts cultured from irradiated skin and effect of TGF-β 1clinical study." Br J Cancer 83(5): 650-654.

Jackman, S. R., O. C. Witard, A. Philp, G. A. Wallis, K. Baar and K. D. Tipton (2017). "Branched-chain amino acid ingestion stimulates muscle myofibrillar protein synthesis following resistance exercise in humans." Front Physiol 8.

Jamshidi, N. and B. Ø. Palsson (2007). "Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets." BMC Syst Biol 1: 26.

Jerby, L., T. Shlomi and E. Ruppin (2010). "Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism." Mol Syst Biol 6: 401.

Jones, D. A., J. M. Round and A. de Haan (2004). Skeletal Muscle from Molecules to Movement: A Textbook of Muscle Physiotherapy for Sport, Exercise and Physiotherapy, Churchill Livingstone.

Jordana, M., A. D. Befus, M. T. Newhouse, J. Bienenstock and J. Gauldie (1988). "Effect of histamine on proliferation of normal human adult lung fibroblasts." Thorax 43(7): 552-558.

Katze, M. G. (2013). Systems Biology, Springer Berlin Heidelberg.

Kesić, S. (2016). "Systems biology, emergence and antireductionism." Saudi J Biol Sci 23(5): 584-591.

Kim, B., W. J. Kim, D. I. Kim and S. Y. Lee (2015). "Applications of genome-scale metabolic network model in metabolic engineering." J Ind Microbiol Biotechnol 42(3): 339-348.

Kim, J.-w., I. Tchernyshyov, G. L. Semenza and C. V. Dang (2006). "HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia." Cell Metabolism 3(3): 177-185.

King, Z. A., A. Drager, A. Ebrahim, N. Sonnenschein, N. E. Lewis and B. Ø. Palsson (2015). "Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways." PLoS Comput Biol 11(8): e1004321.

Kohyama, T., Y. Yamauchi, H. Takizawa, S. Kamitani, S. Kawasaki and T. Nagase (2010). "Histamine stimulates human lung fibroblast migration." Mol Cell Biochem 337(1-2):77-81

Kraggerud, S. M., J. A. Sandvik and E. O. Pettersen (1995). "Regulation of protein synthesis in human cells exposed to extreme hypoxia." Anticancer Res 15(3): 683-686.

Krauss, M., S. Schaller, S. Borchers, R. Findeisen, J. Lippert and L. Kuepfer (2012). "Integrating cellular metabolism into a multiscale whole-body model." PLoS Comput Biol 8(10): e1002750.

Krebs, M., M. Krssak, P. Nowotny, D. Weghuber, S. Gruber, V. Mlynarik, M. Bischof, H. Stingl, C. Fürnsinn, W. Waldhäusl and M. Roden (2001). "Free fatty acids inhibit the glucose-stimulated increase of intramuscular glucose-6-phosphate concentration in humans." J Clin Endocrinol Metab 86(5): 2153-2160.

Kremling, A. (2013). Systems Biology: Mathematical Modeling and Model Analysis, CRC Press.

Kucharzewska, P., H. C. Christianson and M. Belting (2015). "Global profiling of metabolic adaptation to hypoxic stress in human glioblastoma cells." PLoS One 10(1).

Kuchel, P. W. (1998). Schaum's outline of theory and problems of biochemistry, McGraw-Hill.

Lee, S., C. Phalakornkule, M. M. Domach and I. E. Grossmann (2000). "Recursive MILP model for finding all the alternate optima in LP models for metabolic networks." Computers & Chemical Engineering 24(2): 711-716.

Lee, S. C., L. H. Lee, A. J. Pappelis, P. M. BeMiller and J. N. DeMiller (1977). "Dry mass measurements in human diploid fibroblasts." J Histochem Cytochem 25(9): 1105-1106.

Leivonen, S. K., K. Lazaridis, J. Decock, A. Chantry, D. R. Edwards and V. M. Kahari (2013). "TGF-β-elicited induction of tissue inhibitor of metalloproteinases (TIMP)-3 expression in fibroblasts involves complex interplay between Smad3, p38alpha, and ERK1/2." PLoS One 8(2): e57474.

Levick, S. P. and A. Widiapradja (2018). "Mast cells: key contributors to cardiac fibrosis." Int. J. Mol 19(1): 231.

Lewis, N. E., H. Nagarajan and B. Ø. Palsson "Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods." Nat Rev Microbiol 10(4): 291-305.

Lewis, N. E., G. Schramm, A. Bordbar, J. Schellenberger, M. P. Andersen, J. K. Cheng, N. Patel, A. Yee, R. A. Lewis, R. Eils, R. Konig and B. O. Palsson (2010). "Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain." Nat Biotechnol 28(12): 1279-1285.

Liao, S. F., T. Wang and N. Regmi (2015). "Lysine nutrition in swine and the related monogastric animals: muscle protein biosynthesis and beyond." Springerplus 4.

Lieber, R. L. and S. R. Ward (2011). "Skeletal muscle design to meet functional demands." Philos Trans R Soc Lond B Biol Sci 366(1570): 1466-1476.

Lin, L., K. Yamagata, S. Nakayamada, N. Sawamukai, K. Yamaoka, K. Sakata, K. Nakano and Y. Tanaka (2015). "Histamine inhibits differentiation of skin fibroblasts into myofibroblasts." Biochem Biophys Res Commun 463(3): 434-439.

Louard, R. J., E. J. Barrett and R. A. Gelfand (1990). "Effect of infused branched-chain amino acids on muscle and whole-body amino acid metabolism in man." Clin Sci (Lond) 79(5): 457-466.

Louard, R. J., E. J. Barrett and R. A. Gelfand (1995). "Overnight branched-chain amino acid infusion causes sustained suppression of muscle proteolysis." Metabolism 44(4): 424-429.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology 15(12): 550.

Lu, P., K. Takai, V. M. Weaver and Z. Werb (2011). "Extracellular matrix degradation and remodeling in development and disease." Cold Spring Harb Perspect Biol 3(12).

Ma, H., A. Sorokin, A. Mazein, A. Selkov, E. Selkov, O. Demin and I. Goryanin (2007). "The Edinburgh human metabolic network reconstruction and its functional analysis." Mol Syst Biol 3: 135.

Macarak, E. J., M. De la Llera, N. A. Kefalides and B. V. Howard (1979). "Comparison of the intracellular lipids of cultured vascular endothelial cells and fibroblasts." In Vitro 15(12): 936-940.

Mahadevan, R. and B. Ø. Palsson (2005). "Properties of metabolic networks: Structure versus Function." Biophys J 88(1): L07-09.

Mahadevan, R. and C. H. Schilling (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." Metabolic Engineering 5(4): 264-276.

Margolin, A. A. and A. Califano (2007). "Theory and limitations of genetic network inference from microarray data." Ann N Y Acad Sci 1115: 51-72.

Martins Conde Pdo, R., T. Sauter and T. Pfau (2016). "Constraint based modeling going multicellular." Front Mol Biosci 3: 3.

Masoud, G. N. and W. Li (2015). "HIF-1α pathway: role, regulation and intervention for cancer therapy." Acta Pharmaceutica Sinica B 5(5): 378-389.

Mastellar, S. L., R. J. Coleman and K. L. Urschel (2016). "Controlled trial of whole body protein synthesis and plasma amino acid concentrations in yearling horses fed graded amounts of lysine." Vet J 216: 93-100.

May, M. E. and M. G. Buse (1989). "Effects of branched-chain amino acids on protein turnover." Diabetes Metab Rev 5(3): 227-245.

McCall, M. N., K. Uppal, H. A. Jaffee, M. J. Zilliox and R. A. Irizarry (2011). "The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes." Nucleic Acids Res 39(Database issue): D1011-1015.

McKeown, S. R. (2014). "Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response." Br J Radiol 87(1035): 20130676.

McKinnell, I. W. and M. A. Rudnicki (2004). "Molecular mechanisms of muscle atrophy." Cell 119(7): 907-910.

McNurlan, M. A., E. B. Fern and P. J. Garlick (1982). "Failure of leucine to stimulate protein synthesis in vivo." Biochem J 204(3): 831-838.

Meran, S. and R. Steadman (2011). "Fibroblasts and myofibroblasts in renal fibrosis." Int J Exp Pathol 92(3): 158-167.

Mohberg, J. and M. J. Johnson (1963). "Amino acid utilization by 929-L fibroblasts in chemically defined media." J Natl Cancer Inst 31: 611-625.

Morash, A. J., A. O. Kotwica and A. J. Murray (2013). "Tissue-specific changes in fatty acid oxidation in hypoxic heart and skeletal muscle." Am J Physiol Regul Integr Comp Physiol 305(5): R534-541.

Murakami, U. and K. Uchida (1985). "Contents of myofibrillar proteins in cardiac, skeletal, and smooth muscles." J Biochem 98(1): 187-197.

Musutova, M., M. Elkalaf, N. Klubickova, M. Koc, S. Povysil, J. Rambousek, B. Volckaert, F. Duska, M. D. Trinh, M. Kalous, J. Trnka, K. Balusikova, J. Kovar and J. Polak (2018). "The effect of hypoxia and metformin on fatty acid uptake, storage, and oxidation in L6 differentiated myotubes." Frontiers in Endocrinology 9(616).

Nigdelioglu, R., R. B. Hamanaka, A. Y. Meliton, E. O'Leary, L. J. Witt, T. Cho, K. Sun, C. Bonham, D. Wu, P. S. Woods, A. N. Husain, D. Wolfgeher, N. O. Dulin, N. S. Chandel and G. M. Mutlu (2016). "TGF-β promotes *de novo* serine synthesis for collagen production." J Biol Chem 291(53): 27239-27251

Nogiec, C. D. and S. Kasif (2013). "To supplement or not to supplement: A metabolic network framework for human nutritional supplements." PLoS One 8(8): e68751.

Norton, L. E. and D. K. Layman (2006). "Leucine regulates translation initiation of protein synthesis in skeletal muscle after exercise." J Nutr 136(2): 533S-537S.

Novick, P., B. C. Osmond and D. Botstein (1989). "Suppressors of yeast actin mutations." Genetics 121(4): 659-674.

O'Brien, E. J., J. M. Monk and B. Ø. Palsson (2015). "Using genome-scale models to predict Biological Capabilities." Cell 161(5): 971-987.

O'Bryan, K. R., T. M. Doering, R. W. Morton, V. G. Coffey, S. M. Phillips and G. R. Cox (2019). "Do multi-ingredient protein supplements augment resistance training-induced gains in skeletal muscle mass and strength? A systematic review and meta-analysis of 35 trials." Br J Sports Med.

Oberhardt, M. A., B. Ø. Palsson and J. A. Papin (2009). "Applications of genome-scale metabolic reconstructions." Mol Syst Biol 5: 320.

Oh, Y. K., B. Ø. Palsson, S. M. Park, C. H. Schilling and R. Mahadevan (2007). "Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data." J Biol Chem 282(39): 28791-28799.

Orth, J. D., I. Thiele and B. Ø. Palsson (2010). "What is flux balance analysis?" Nat Biotechnol 28(3): 245-248.

Ottenheijm, C. A., H. Granzier and S. Labeit (2012). "The sarcomeric protein nebulin: another multifunctional giant in charge of muscle strength optimization." Front Physiol 3: 37.

Pacheco, M. P., E. John, T. Kaoma, M. Heinäniemi, N. Nicot, L. Vallar, J.-L. Bueb, L. Sinkkonen and T. Sauter (2015). "Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network." BMC Genomics 16(1): 809.

Palsson, B. Ø. (2000). "The challenges of *in silico* biology." Nat Biotechnol 18: 1147.

B. Ø. Palsson (2015). Systems Biology, Cambridge University Press.

Pasiakos, S. M., C. E. Berryman, C. T. Carrigan, A. J. Young and J. W. Carbone (2017). "Muscle protein turnover and the molecular regulation of muscle mass during hypoxia." Med Sci Sports Exerc 49(7): 1340-1350.

Passey, S. L., M. J. Hansen, S. Bozinovski, C. F. McDonald, A. E. Holland and R. Vlahos (2016). "Emerging therapies for the treatment of skeletal muscle wasting in chronic obstructive pulmonary disease." Pharmacology & Therapeutics 166: 56-70.

Petrov, V. V., R. H. Fagard and P. J. Lijnen (2002). "Stimulation of collagen production by transforming growth factor-β 1 during differentiation of cardiac fibroblasts to myofibroblasts." Hypertension 39(2): 258-263.

Pfau, T., N. Christian and O. Ebenhöh (2011). "Systems approaches to modelling pathways and networks." Briefings in Functional Genomics 10(5): 266-279.

Pfau, T., M. P. Pacheco and T. Sauter (2016). "Towards improved genome-scale metabolic network reconstructions: unification, transcript specificity and beyond." Brief Bioinform 17(6): 1060-1069.

Pfeiffer, T., O. S. Soyer and S. Bonhoeffer (2005). "The evolution of connectivity in metabolic networks." PLoS Biology 3(7): e228.

Phillips, S. M. (2014). "A brief review of critical processes in exercise-induced muscular hypertrophy." Sports Med 44(Suppl 1): 71-77.

Pozo, C., G. Guillen-Gosalbez, A. Sorribas and L. Jimenez (2012). "Identifying the preferred subset of enzymatic profiles in nonlinear kinetic metabolic models via multiobjective global optimization and Pareto filters." PLoS One 7(9): e43487.

Price, N. D., J. L. Reed and B. Ø. Palsson (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." Nat Rev Microbiol 2: 886.

Raghu, G., H. R. Collard, J. J. Egan, F. J. Martinez, J. Behr, K. K. Brown, T. V. Colby, J. F. Cordier, K. R. Flaherty, J. A. Lasky, D. A. Lynch, J. H. Ryu, J. J. Swigris, A. U. Wells, J. Ancochea, D. Bouros, C. Carvalho, U. Costabel, M. Ebina, D. M. Hansell, T. Johkoh, D. S. Kim, T. E. King, Y. Kondoh, J. Myers, N. L. Müller, A. G. Nicholson, L. Richeldi, M. Selman, R. F. Dudden, B. S. Griss, S. L. Protzko and H. J. Schünemann (2011). "An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis." Am J Respir Crit Care Med 183(6): 788-824.

Ramakrishna, R., J. S. Edwards, A. McCulloch and B. Ø. Palsson (2001). "Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints." Am J Physiol Regul Integr Comp Physiol 280(3): R695-R704.

Randle, P. J., P. B. Garland, C. N. Hales and E. A. Newsholme (1963). "The glucose fatty-acid cycle. Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus." Lancet 1(7285): 785-789.

Reed, J. L. and B. Ø. Palsson (2004). "Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network States." Genome Res 14(9): 1797-1805.

Robinson, J. L. and J. Nielsen (2016). "Integrative analysis of human omics data using biomolecular networks." Molecular BioSystems 12(10): 2953-2964.

Roden, M. (2004). "How free fatty acids inhibit glucose utilization in human skeletal muscle." News Physiol Sci 19: 92-96.

Everest. II: Nutrition and body composition." J Appl Physiol 65(6): 2545-2551.

Rose, M. S., C. S. Houston, C. S. Fulco, G. Coates, J. R. Sutton and A. Cymerman (1988). "Operation Everest. II: Nutrition and body composition." J Appl Physiol (1985) 65(6): 2545-2551.

Rugen, M., A. Bockmayr, J. Legrand and G. Cogne (2012). "Network reduction in metabolic pathway analysis: elucidation of the key pathways involved in the photoautotrophic growth of the green alga *Chlamydomonas reinhardtii*." Metab Eng 14(4): 458-467.

Salceda, S. and J. Caro (1997). "Hypoxia-inducible Factor 1α (HIF-1α) Protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions: Its stabilization by hypoxia depends on redox-induced changes." J Biol Chem 272(36): 22642-22647.

Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards and B. Ø. Palsson (2002). "Genome-scale metabolic model of *Helicobacter pylori* 26695." J Bacteriol 184(16): 4582-4593.

Schilling, C. H., J. S. Edwards, D. Letscher and B. Ø. Palsson (2000). "Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems." Biotechnol Bioeng 71(4): 286-306.

Schilling, C. H. and B. Ø. Palsson (2000). "Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis." J Theor Biol 203(3): 249-283.

Schilling, C. H., S. Schuster, B. Ø. Palsson and R. Heinrich (1999). "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era." Biotechnology Progress 15(3): 296-303.

Schock, E. N., C. F. Chang, I. A. Youngworth, M. G. Davey, M. E. Delany and S. A. Brugmann (2016). "Utilizing the chicken as an animal model for human craniofacial ciliopathies." Dev Biol 415(2): 326-337.

Schwender, J. (2008). "Metabolic flux analysis as a tool in metabolic engineering of plants." Curr Opin Biotechnol 19(2): 131-137.

Segrè, D., D. Vitkup and G. M. Church (2002). "Analysis of optimality in natural and perturbed metabolic networks." Proc Natl Acad Sci U S A 99(23): 15112-15117.

Selvarajah, B., I. Azuelos, M. Platé, D. Guillotin, E. J. Forty, G. Contento, H. V. Woodcock, M. Redding, A. Taylor, G. Brunori, P. F. Durrenberger, R. Ronzoni, A. D. Blanchard, P. F. Mercer, D. Anastasiou and R. C. Chambers (2019). "mTORC1 amplifies the ATF4-dependent de novo serine-glycine pathway to supply glycine during TGF-β(1)-induced collagen biosynthesis." Sci Signal 12(582).

Semenza, G. L., P. H. Roth, H. M. Fang and G. L. Wang (1994). "Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1." J Biol Chem 269(38): 23757-23763.

Shah, J. B. (2011). "The history of wound care." J Am Col Certif Wound Spec 3(3): 65-66.

Shah, M., D. M. Foreman and M. W. J. Ferguson (1992). "Control of scarring in adult wounds by neutralising antibody to transforming growth factor β." The Lancet 339(8787): 213-214.

Shlomi, T., M. N. Cabili and E. Ruppin (2009). "Predicting metabolic biomarkers of human inborn errors of metabolism." Mol Syst Biol 5: 263.

Shlomi, T., Y. Eisenberg, R. Sharan and E. Ruppin (2007). "A genome-scale computational study of the interplay between transcriptional regulation and metabolism." Mol Syst Biol 3: 101.

Smith, K., N. Reynolds, S. Downie, A. Patel and M. J. Rennie (1998). "Effects of flooding amino acids on incorporation of labeled amino acids into human muscle protein." Am J Physiol 275(1 Pt 1): E73-78.

Solaini, G., A. Baracca, G. Lenaz and G. Sgarbi (2010). "Hypoxia and mitochondrial oxidative metabolism." Biochimica et Biophysica Acta (BBA) - Bioenergetics 1797(6): 1171-1177.

Srivastava, S. P., J. Li, M. Kitada, H. Fujita, Y. Yamada, J. E. Goodwin, K. Kanasaki and D. Koya (2018). "SIRT3 deficiency leads to induction of abnormal glycolysis in diabetic kidney with fibrosis." Cell Death & Disease 9(10): 997.

Stepanenko, A. A. and V. V. Dmitrenko (2015). "HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution." Gene 569(2): 182-190.

Subramanian, M., P. P. Kuang, L. Wei, D. C. Rishikof, H. Liu and R. H. Goldstein (2006). "Modulation of amino acid uptake by TGF-β in lung myofibroblasts." J Cell Biochem 99(1): 71-78.

Swainston, N., K. Smallbone, H. Hefzi, P. D. Dobson, J. Brewer, M. Hanscho, D. C. Zielinski, K. S. Ang, N. J. Gardiner, J. M. Gutierrez, S. Kyriakopoulos, M. Lakshmanan, S. Li, J. K. Liu, V. S. Martinez, C. A. Orellana, L. E. Quek, A. Thomas, J. Zanghellini, N. Borth, D. Y. Lee, L. K. Nielsen, D. B. Kell, N. E. Lewis and P. Mendes (2016). "Recon 2.2: from reconstruction to model of human metabolism." Metabolomics 12: 109.

Tamm, E. R., A. Siegner, A. Baur and E. LÜTjen-Drecoll (1996). "Transforming growth factor-β1 induces α-Smooth Muscle-Actin expression in cultured human and monkey trabecular meshwork." Experimental Eye Research 62(4): 389-398.

Tate, J. G., S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell and S. A. Forbes (2018). "COSMIC: the catalogue of somatic mutations in cancer." Nucleic Acids Res 47(D1): D941-D947.

Tello-Ruiz, M. K., S. Naithani, J. C. Stein, P. Gupta, M. Campbell, A. Olson, S. Wei, J. Preece, M. J. Geniza, Y. Jiao, Y. K. Lee, B. Wang, J. Mulvaney, K. Chougule, J. Elser, N. Al-Bader, S. Kumari, J. Thomason, V. Kumar, D. M. Bolser, G. Naamati, E. Tapanari, N. Fonseca, L. Huerta, H. Iqbal, M. Keays, A. Munoz-Pomer Fuentes, A. Tang, A. Fabregat, P. D'Eustachio, J. Weiser, L. D. Stein, R. Petryszak, I. Papatheodorou, P. J. Kersey, P. Lockhart, C. Taylor, P. Jaiswal and D. Ware (2018). "Gramene 2018: unifying comparative genomics and pathway resources for plant research." Nucleic Acids Res **46**(D1): D1181-d1189.

Thiele, I. and B. Ø. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nat Protoc 5: 93.

Thiele, I., N. Swainston, R. M. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, Sr., M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. H. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes and B. Ø. Palsson (2013). "A community-driven global reconstruction of human metabolism." Nat Biotechnol 31(5): 419-425.

Thiele, I., N. Vlassis and R. M. Fleming (2014). "fastGapFill: efficient gap filling in metabolic networks." Bioinformatics 30(17): 2529-2531.

Tipton, K. D., A. A. Ferrando, S. M. Phillips, D. Doyle, Jr. and R. R. Wolfe (1999). "Postexercise net protein synthesis in human muscle from orally administered amino acids." Am J Physiol 276(4 Pt 1): E628-634.

Tipton, K. D. and R. R. Wolfe (2001). "Exercise, protein metabolism, and muscle growth." Int J Sport Nutr Exerc Metab 11(1): 109-132.

Tobalina, L., J. Pey, A. Rezola and F. J. Planes (2016). "Assessment of FBA based gene essentiality analysis in cancer with a fast context-specific network reconstruction method." PLoS One 11(5): e0154583.

Tomasek, J. J., G. Gabbiani, B. Hinz, C. Chaponnier and R. A. Brown (2002). "Myofibroblasts and mechano-regulation of connective tissue remodelling." Nat Rev Mol Cell Bio 3: 349.

Trewavas, A. (2006). "A Brief History of Systems Biology: “Every object that biology studies is a system of systems.” Francois Jacob (1974)." Plant Cell 18(10): 2420-2430.

Turcotte, L. P., C. Petry, B. Kiens and E. A. Richter (1998). "Contraction-induced increase in Vmax of palmitate uptake and oxidation in perfused skeletal muscle." J Appl Physiol (1985) 84(5): 1788-1794.

Uhlen, M., L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L.

Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen and F. Ponten (2015). "Proteomics. Tissue-based map of the human proteome." Science 347(6220): 1260419.

Umapathysivam, K., J. J. Hopwood and P. J. Meikle (2005). "Correlation of acid alpha-glucosidase and glycogen content in skin fibroblasts with age of onset in Pompe disease." Clin Chim Acta 361(1-2): 191-198.

Vaidyanathan, S., G. G. Harrigan and R. Goodacre (2006). Metabolome Analyses: Strategies for Systems Biology, Springer US.

Verrecchia, F. and A. Mauviel (2007). "Transforming growth factor-β and fibrosis." World J Gastroenterol 13(22): 3056-3062.

Vlassis, N., M. P. Pacheco and T. Sauter (2014). "Fast reconstruction of compact context-specific metabolic network models." PLoS Computational Biology 10(1): e1003424.

Vo, T. D., H. J. Greenberg and B. Ø. Palsson (2004). "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data." J Biol Chem 279(38): 39532-39540.

Vo, T. D., W. N. Paul Lee and B. Ø. Palsson (2007). "Systems analysis of energy metabolism elucidates the affected respiratory chain complex in Leigh's syndrome." Mol Genet Metab 91(1): 15-22.

Volk, S. W., Y. Wang, E. A. Mauldin, K. W. Liechty and S. L. Adams (2011). "Diminished type III collagen promotes myofibroblast differentiation and increases scar deposition in cutaneous wound healing." Cells Tissues Organs 194(1): 25-37.

von Bertalanffy, L. (2003). General System Theory: Foundations, Development, Applications, G. Braziller.

Wang, K. (1996). "Titin/connectin and nebulin: Giant protein rulers of muscle structure and function." Advances in Biophysics 33: 123-134.

Wang, X., S. Qiao, Y. Yin, L. Yue, Z. Wang and G. Wu (2007). "A deficiency or excess of dietary threonine reduces protein synthesis in jejunum and skeletal muscle of young pigs." J Nutr 137(6): 1442-1446.

Wang, Y., J. A. Eddy and N. D. Price (2012). "Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE." BMC Syst Biol 6: 153.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet 10: 57.

Warburg, O. (1956). "On the origin of cancer cells." Science 123(3191): 309-314.

Webber, J., R. H. Jenkins, S. Meran, A. Phillips and R. Steadman (2009). "Modulation of TGF-β 1-dependent myofibroblast differentiation by hyaluronan." Am J Pathol 175(1): 148-160.

Weinstein, D., J. B. Marsh, M. Catherine Glick and L. Warren (1969). "Membranes of animal cells. IV. Lipids of the L cell and its surface membrane." J Biol Chem 244(15): 4103-11

Wendt, M. K. and W. P. Schiemann (2009). "Therapeutic targeting of the focal adhesion complex prevents oncogenic TGF-β signaling and metastasis." Breast Cancer Res 11(5): R68.

Westerhoff, H. V. and B. Ø. Palsson (2004). "The evolution of molecular biology into systems biology." Nat. Biotechnol 22: 1249.

Wilborn, C. D. and D. S. Willoughby (2004). "The role of dietary protein intake and resistance training on myosin heavy chain expression." J Int Soc Sports Nutr 1(2): 27-34.

Williams, P. E. and G. Goldspink (1971). "Longitudinal growth of striated muscle fibres." J Cell Sci 9(3): 751-767.

Wolfe, R. R. (2017). "Branched-chain amino acids and muscle protein synthesis in humans: myth or reality?" J Int Soc Sports Nutr 14.

Wolfenden, R. and M. J. Snider (2001). "The depth of chemical time and the power of enzymes as catalysts." Acc Chem Res 34(12): 938-945.

Xie, N., Z. Tan, S. Banerjee, H. Cui, J. Ge, R. M. Liu, K. Bernard, V. J. Thannickal and G. Liu (2015). "Glycolytic reprogramming in myofibroblast differentiation and lung fibrosis." Am J Respir Crit Care Med 192(12): 1462-1474.

Yates, B., B. Braschi, K. A. Gray, R. L. Seal, S. Tweedie and E. A. Bruford (2017). "Genenames.org: the HGNC and VGNC resources in 2017." Nucleic Acids Res **45**(D1): D619-d625.

Yokozeki, M., K. Moriyama, H. Shimokawa and T. Kuroda (1997). "Transforming growth factor-β modulates myofibroblastic phenotype of rat palatal fibroblasts *in vitro*." Exp. Cell Res 231(2): 328-336

Zank, D. C., M. Bueno, A. L. Mora and M. Rojas (2018). "Idiopathic Pulmonary Fibrosis: aging, mitochondrial dysfunction, and cellular bioenergetics." Frontiers in Medicine 5(10).

Zhang, D., Y. Wang, Z. Shi, J. Liu, P. Sun, X. Hou, J. Zhang, S. Zhao, Binhua P. Zhou and J. Mi (2015). "Metabolic Reprogramming of cancer-associated fibroblasts by IDH3α Downregulation." Cell Reports 10(8): 1335-1348.

Zhao, S., W. P. Fung-Leung, A. Bittner, K. Ngo and X. Liu (2014). "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells." PLoS One 9(1): e78644.

Zierath, J. R., T.-S. Tsao, A. E. Stenbit, J. W. Ryder, D. Galuska and M. J. Charron (1998). "Restoration of Hypoxia-stimulated Glucose Uptake in GLUT4-deficient Muscles by Muscle-specific GLUT4 Transgenic Complementation." J Biol Chem 273(33): 20910-20915.

Zion Market Research (2017). "Dietary supplements market by ingredients (Botanicals, Vitamins, Minerals, Amino Acids, Enzymes) for additional supplements, medicinal supplements and sports nutrition applications - Global industry perspective, comprehensive analysis and forecast, 2016 – 2022" https://www.zionmarketresearch.com/report/dietary-supplements-market

List of URLs

www.uniprot.org

www.bioconductor.org

www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836

https://github.com/andreidob91/cobramodels